Vol. XII Part II The British Journal of Statistical Psychology November 1959

THE MEASUREMENT OF INTERESTS BY VERBAL METHODS

By E. A. Peel
University of Birmingham

The following paper describes verbal tests of information and vocabulary designed to measure the school and leisure interests of pupils. To meet the common criticism that such tests measure attainment rather than interest a special technique has been devised, that of forcing a choice upon the testee.

In the first form of the test a choice has to be made between two equally difficult items from two different fields (e.g., practical and academic, scientific and non-scientific, or the like). Evidence is offered for its validity and reliability. In the second form a suitable list of homonyms is drawn up (e.g., words having both a practical and an academic meaning), and the testee is asked to give his interpretation of each.

The paper concludes with a theoretical discussion of the relation of attainment to ability and interest; and it is suggested that ability and interest should be combined multiplicatively to produce attainment rather than additively as in most factorial studies.

I. Introduction

For some years my research students and I have been carrying out investigations on the possibility of measuring pupils' interests in various fields by means of verbal tests. This work began in 1947 when Dr. C. M. Lambert, H.M.I., and I collaborated [1, 2, 3] on work she was then carrying out as Tests and Records Officer to the Northumberland Local Education Authority. At that time we were primarily concerned with the use that could be made, in allocating pupils to secondary schools at the age of 11+, of their interests in academic or practical subjects. The tests developed showed such promise that the method was applied to other fields of interest both in and out of school. Angus [4] used the same technique to distinguish between the interests of middle school pupils of grammar schools in scientific and non-scientific subjects, and to study the feasibility of classifying them into science and arts streams. Pennington [5, 6] used it to investigate the secondary schoolboys' leisure time interests. More recently the original Lambert–Peel interests have been adopted in follow-up studies in secondary schools in the West Riding [7] and tentative investigations on the use of homonyms as a test of bias of interest, previously started in Durham, were extended as a part of the inquiry [8, 9]. The use of vocabulary items was developed by Burroughs [10] for assessing the interests of teacher applicants. Dr. T. W. G. Miller has applied the technique to assess the leisure interests of boys in comprehensive, grammar, and modern schools in connection with an inquiry into interests, attitudes, and morale in comprehensive schools [11].

My own attention has lately turned to the psychometric problems thus raised, and to questions regarding the relation of interest to motivation and other modes of reinforcement. The object of the present paper is to discuss this work in order to ventilate a topic of special importance in educational and social psychology. The methods used raise interesting psychometric points, and suggest hypotheses about the relations between fundamental psychological categories. The material will therefore be partly empirical and partly speculative.

II. Techniques

Verbal and informational methods of measuring interests are based on the assumption that, if an interest is real and active, it will result in the acquisition of information about the field or topic of interest. The chief difficulty is to ensure that tests of verbal information or vocabulary assess *interest* rather than *ability* or *attainment*. The theoretical aspects of this problem will be deferred to section V; and I shall first describe how we tried to deal with it in practical situations.

From a practical standpoint there are three questions to be considered: the sample of items to be used, the method of presenting them, and the sample of persons tested. Suppose we wish to test interest in mechanical craftsmanship; we could prepare a list of terms used—e.g. *tenon saw*, *mitre joint*, *soldering iron*, *wiped joint*, *calipers*, *spirit level*; and ask the testees to explain the meaning of as many items as possible. Such a list, however, would hardly differentiate interest from ability and attainment; and Fryer declares ([12], p. 290) that nothing different from ability can be measured by tests of information. Dr. Lambert's original lists revealed the same tendency; and similar criticisms could be made of the tests devised by Wyman [14], and of the selective vocabulary test devised by Slater. The latter is composed of a mixed list of words of feminine and masculine interest. Here we encounter the further difficulty, namely, how to eliminate differences resulting from ability or education. An able and well educated person of the 'masculine' type may be able to cope far more easily with 'feminine' items than an average person of the 'feminine' type. In this case, however, the difficulty might be largely overcome by assessing the ratio of masculine to feminine responses instead of the totals.

To meet such difficulties the first requirement is to compel the testee to choose between items representing two opposite fields of interest—in our case practical and academic interests respectively. These were plainly relevant to the problem of selecting pupils for education in technical or grammar schools. In each set the items from which the choice was made had to be of equal difficulty; and only a fixed number, three out of six, had to be answered. The six items consisted of words, small drawings, or pieces of information, which were set out in random order, and related either to practical or to literary topics. In the case of the practical items the answers could be not only in words but by means of drawings or diagrams. The following specimens are taken from the vocabulary

test[1]: I. Brace and bit (P), parable, fuselage (P), duet, furlong, chisel (P). II. Lathe (P), treaty, spirit level (P), pressure gauge (P), plateau, peninsula. III. Geologist, renaissance, crankshaft (P), joist (P), engine coupling rod (P), serf. (The letter P is inserted to show the reader which are the 'practical' items.) The test included seven such sets, and was described as 'a meaning of words' test. The answers could range from 21 practical and 0 academic responses to 21 academic and 0 practical responses, according to the interests of each testee. The essential requirement is that the six items making up any sub-set should be of equal difficulty. This is met by collecting a large number of items, mixing them at random and giving them to a group of pupils in such a way that *all* the items shall be attempted. The 'difficulty' value of each is then expressed as percentages; and, after throwing out invalid items, sets of six (three practical and three academic) are made up, so that the most difficult and the most easy in each set do not differ by more than 10 per cent in difficulty. Very easy and very hard items are rejected; those chosen lie within the range of 80 to 20 per cent. The reason is that the test proves most effective as a measure of interest (in distinction from ability and attainment) when exactly 21 correct responses are given by all the testees, and the assessment turns, not on the number but on the composition of these 21 responses. Indeed in later experiments (those of Miller) the range of difficulty was further narrowed to between 70 and 30 per cent.

With seven sets of six items the test can now be administered. Since in practice it is difficult to obtain exactly 21 correct responses from every pupil, it becomes necessary to consider the best way of scoring the test, so as to allow for the variations. First, we obtain the two scores for practical (P) and the academic (A) responses respectively. Ideally $P+A$ should equal 21; and, were this always the case, we could express the bias in interest by the difference, $P-A$. But $P+A$ varies; and this variation can presumably reflect a variation in ability or attainment that we have failed to eliminate, it therefore becomes necessary to compensate for it. The formulae used for this purpose were (i) $(P-A)/(P+A)\times 100$, (ii) $100P/(P+A)$, and (iii) $100A/(P+A)$.

[1] This vocabulary test forms Part II of the Lambert–Peel General Information Test; Part I is made up of eight sets of six items of *information*, three Practical, three Academic, of which the following are specimens of the practical and academic type respectively.

1. Practical examples:
   (*a*) Here are names of parts of a bicycle. Underline the one which contains ball bearings: *lamp, handlebars, pedal, brake block.*
   (*b*) Imagine you are to assemble an electric iron from the parts A, B, C, D, E illustrated below. Show by using the letters the order in which you would put the parts together. (Drawings of A, B, C, D, E are given.)
2. Academic examples:
   (*a*) Here are names from ancient history. Underline the name of the man famed for his journey over the Alps: Hannibal, Archimedes, Brutus, Cicero, Romulus.
   (*b*) What is a constellation? Underline the correct answer: a group of meteors, the innermost planets of the solar system, a group of stars easily recognized by the naked eye and forming a definite pattern, a comet which has lost its tail.

The technique used in constructing this test was exactly the same as for the vocabulary test.

The correction achieved by dividing by P + A, however, is not satisfactory for pupils who answer only a few items; it tends to exaggerate the bias of interest. This may be avoided by refusing to accept low P + A scores as a measure of interest. In practice the difficulty is not so great as might be imagined, since a test of practical versus academic interests is required only for the top 30 per cent of the age-group. The test is least useful where it is least needed. Thus we have virtually met all the three requirements set out above, viz., (i) the equating of the absolute difficulty of the six items making up any sub-test, (ii) the introduction of an element of choice to reveal the bias of interest, and (iii) the elimination of large differences in ability and attainment between individuals tested.

III. Results

Total scores. Let us first consider how far the test results fulfil the condition that the (P + A) scores should not diverge too widely. Pennington [5] in his studies of Grammar and Modern School groups, contrasted intellectual (I) and active (A) interests, intellectual (I) and passive (P) interests, and active (A) and passive (P) interests—the maximum possible in each case being 24. Table I shows the number (n), mean (m), and standard deviation (s) for each group.

Table I. Means and Standard Deviations for Total Scores (Pennington)

		I+A		I+P		A+P	
Group	n	m	s	m	s	m	s
Grammar	84	15·46	2·92	16·92	2·56	15·52	2·54
Modern							
Stream 1	36	14·82	3·24	15·78	2·98	14·66	2·84
Stream 2	31	11·24	2·48	12·92	3·50	13·18	3·34
Stream 3	27	9·58	2·46	10·54	4·20	10·38	2·90
Stream 4	24	6·58	2·94	7·16	4·22	6·50	3·26

Even with the grammar school group, which was fairly uniform intellectually, the total scores are far from constant; still their dispersion is not excessive (under 20 per cent of the mean). Further, in the secondary modern school as we move from the first stream to the fourth there is an appreciable decline in the scores; and Pennington did in fact exclude all whose scores fell below six. Cureton and Potgieter found similar results, but included in their inquiry only those pupils who had a constant P + A total of 24, the maximum possible being 42: in this way they held ability and attainment constant. Miller used a vocabulary test in which he opposed active and passive leisure interests. The maximum score was 60. His results are shown in Table II. For the grammar

Table II. Means and Standard Deviations for Total Scores (Miller)

	n	m	s
Combined Grammar	173	46·7	5·45
Combined Comprehensive (Grammar)	129	45·0	6·67
Combined Modern	223	37·11	7·84
Combined Comprehensive (Modern)	225	37·65	7·34

schools the means are much alike; but there is an appreciable scatter (over 10 per cent of the means). The modern groups show a wider dispersion, and there is a significant drop in their means. For the most part the scores for grammar groups lay between 35 and 55 marks, and for the modern between 20 and 50. In an inquiry now in progress, a student investigating the arithmetic versus non-arithmetic interests of 42 secondary modern children has obtained the following results: mean 13·65, standard deviation 2·49; here again there is an appreciable spread (22 per cent of mean). But the whole problem of equating total scores calls for further study; other ways of solving it might well be examined, e.g. using only items of medium difficulty.

Validity. To measure the validity of interest tests directly is far from easy. Direct methods require a comparison of the results from a verbal testing of interests with those from questionnaires and rating scales. This was undertaken by Angus [4]. He used several methods of assessing interest; and found a group factor clearly demarcating the interest tests from other types of measurement. Table III shows the correlations between the teacher's estimate of each pupil's interest in science and the assessments obtained from an interest questionnaire and from vocabulary and information tests of the type described.

Table III. Validity of Tests

	Vocabulary	Information	Interest Questionnaire
Correlation	0·647	0·803	0·387
Standard error	0·053	0·031	0·078

Indirect methods of validation can be applied much more readily. These depend on the assumption that interest, as well as ability, is an essential factor contributing to a pupil's attainments. Thus we should expect a pupil's interests to be revealed by his unusually higher attainments in subjects that interest him. In the research carried out by Dr. Lambert and myself an indirect method was used to discover whether the practical and non-practical tests were positively related to ability tests with similar material ([2], p. 46). We correlated the practical interest sources, obtained from the two verbal tests described above, with Alexander's Performance Scale, a Moray House Verbal Intelligence test, and a practical test consisting of selecting and assembling blocks of wood to form shapes specified by half-scale models or perspective drawings. The correlations between the assessment of practical interest and the results of the other tests (with standard errors) are shown in Table IV. Since the interest

Table IV. Test Correlations

	Correlations	S.E.
Moray House Verbal Intelligence	−0·261	0·071
Alexander Performance Scale	0·249	0·071
Peel Practical Performance	0·233	0·072

tests were designed to oppose practical and academic interest, these coefficients may be regarded as measuring the validity of the tests employed for this purpose.

Cureton [7] compared scores obtained by 102 secondary modern school pupils for practical (P) and academic (A) interests with assessments of their verbal intelligence (I) and school attainments in English (E), mathematics (M), French (F), science (S), art (a), and woodwork (W), tested at the end of their first year. The correlation matrix was analysed by the centroid method. The

TABLE V. INTERCORRELATIONS FOR TESTS

		A	I	E	M	F	S	a	W
P	—	0·023	0·199	0·177	0·155	−0·272	0·332	0·342	0·520
A	0·023	—	0·323	0·440	0·140	0·239	0·267	0·207	0·156
I	0·199	0·323	—	0·377	0·179	0·196	0·239	0·221	−0·001
E	0·177	0·440	0·377	—	0·720	0·341	0·564	0·419	0·437
M	0·155	0·140	0·179	0·720	—	0·464	0·632	0·348	0·332
F	−0·272	0·239	0·196	0·341	0·464	—	0·370	0·054	0·097
S	0·332	0·267	0·239	0·564	0·632	0·370	—	0·459	0·368
a	0·342	0·207	0·221	0·419	0·348	0·054	0·459	—	0·407
W	0·520	0·156	−0·001	0·437	0·332	0·097	0·368	0·407	—
Factor I	0·394	0·441	0·416	0·827	0·726	0·385	0·762	0·575	0·543
Factor II	0·652	−0·233	0·100	−0·251	−0·285	−0·558	0·109	0·267	0·260
Factor I	0·00	0·50	0·30	0·83	0·76	0·62	0·58	0·35	0·32
Factor II	0·76	0·03	0·31	0·22	0·13	−0·22	0·50	0·53	0·51

Note. The smallest correlation which is at least three times its standard error is 0·275.

saturations for the first two factors before and after rotation, are shown at the foot of the table.

The two rotated factors are clearly related to the two main interests, P and A; and the size and signs of their saturations for the different school subjects vary as we might expect. These results furnish evidence for the validity of the tests; and are further corroborated by the very similar results obtained by Dr. Lambert in validity in her tests in a school in Northumberland [2].

With verbal tests of leisure time interests a different form of indirect validation may be employed. Pennington's tests [6] were given to 14-year-old boys in a town grammar school (with A and B streams), a rural grammar school, and a secondary modern school with four streams. The following comparisons can therefore be made: Grammar versus Modern, Higher versus Lower Streams, Rural Grammar versus Town Grammar. His results are shown in Table VI. Interest is measured by the formula $(I-A)/(I+A)\times 100$ in favour of the first member of each pair; I and A are total I and A scores. Almost every entry in the table has a significance in terms of the schools chosen. We see how the grammar school pupils possess the more intellectual interests, and how such interests decrease in the modern school as we pass from the A stream to the D, while their passive interests increase. The influence of town as opposed to

country life is plainly brought out. Doubtless the passive interests of the town grammar school and the more active interest of the country school reflect the leisure opportunities available in the two contrasted areas.

TABLE VI. ASSESSMENT OF INTERESTS IN VARIOUS TYPES OF SCHOOL

School	Class	No.	Intellectual versus Active	Intellectual versus Passive	Active versus Passive
Urban Grammar	IVa	28	+12·0	− 3·2	+ 1·4
	IVb	30	+ 3·7	0·0	+ 3·3
Rural Grammar		26	− 1·0	+ 6·2	+ 8·1
Modern	Stream				
	A	36	− 8·7	− 7·0	−13·1
	B	31	−16·2	−19·3	−21·8
	C	27	−25·0	−43·7	−20·0
	D	24	−11·4	−48·0	−26·1

The most recent evidence for the validity of such tests is supplied by Miller. He investigated such intangible factors as the tone, morale, attitudes, and interests of pupils in comprehensive schools [11]. To discover the nature of their interests he contrasted active interests with passive; and compared four groups of boys—those attending grammar schools, those attending modern schools, a comprehensive ' grammar ' group made up of the abler pupils likely to go on to G.C.E.(O.), and a comprehensive ' modern ' group composed of those of lower ability likely to leave at 15. The results are shown in Table VII. Bias is calculated by the formula $(A-P)/(A+P) \times 100$. The evidence for validity arises from the differences between the mean scores. The increasing bias of interest towards passive interests is obvious as one considers in turn first the grammar groups, next the comprehensive modern group, and finally the modern school.

TABLE VII. MEAN SCORES, RANGE OF SCORES, AND DIFFERENCES IN A TEST OF LEISURE INTERESTS

	No.	Mean Scores	Range	Differences	Sig.
Grammar	174	− 3·66	−71 to 65	GS–CG, 3·44	no
Comprehensive (G)	132	− 0·22	−56 to 85	GS–MS, −21·02	0·01
Comprehensive (M)	217	−14·63	−84 to 72	CG–CM, 14·41	0·01
Modern	224	−24·68	−84 to 56	CM–MS, −10·05	0·01

Last of all, we may cite the correlations obtained by Evans in an unpublished research on secondary school pupils' interests in arithmetic [16]: the correlation of the arithmetic vocabulary test with the tests of arithmetical attainments and of verbal ability amounted to 0·436 and 0·254 respectively.

Reliability. To assess reliability we have the results of Pennington and Miller who used the split-half method. Pennington's figures were obtained from tests whose items had not been validated applied to 198 children; they are shown in Table VIII. Except perhaps for the first test, the coefficients appear

TABLE VIII. RELIABILITY OF INTEREST TESTS

	Coefficient	S.F.
Intellectual v. Active	0·440	0·058
Passive v. Intellectual	0·612	0·045
Active v. Passive	0·669	0·040

distinctly promising, particularly when we bear in mind the shortness of the tests and the possible inclusion of invalid items.

Miller validated his items against the whole test, and rejected invalid items before re-assembling the test material for further use. His test was also longer than Pennington's. With a group of 486 pupils he obtained a split-half reliability of 0·835; and this was raised to 0·910 on using the Spearman–Brown ' prophesy formula '. Here the interest test compares quite well with the most reliable verbal tests in general use.

IV. TESTS BASED ON HOMONYMS

As an alternative method of testing interests we may investigate the differences in interpretation given to single words whose meanings are ambiguous. Thus the word ' brace ' might evoke such different responses as ' a pair ', ' a support ', or ' a joiner's tool for drilling holes '. This suggests that, if we compile a list of homonyms, such as *vice*, *strut*, *tender*, with practical and non-practical meanings, and present it as a vocabulary test, we may get a set of answers which vary with persons predominantly practical, with others predominantly non-practical. Such a test would be essentially a projective test.

Since the first applications of this device were carried out, references to the use of a similar technique by Thurstone [17], Goodenough (*ap.* [18]), and Foley and MacMillan [19] have come to my notice. Thurstone was interested in the social-mindedness of his subjects, and collected sets of homonyms, each of which possessed human or social meaning and a literal or physical meaning. Foley and MacMillan presented homophones, each of which possessed a legal, a medical, and a neutral meaning (e.g. *constitution*, *confinement*, *discharge*) and gave them to legal, medical, and non-professional students. The responses showed that the test gave a valid discrimination between the two professional groups.

In constructing a test to differentiate between practical and non-practical interests in school children, we assume that, if a word has two opposite meanings, the more practically-minded child is more likely to define the word in its practical sense (the term ' practically-minded ' may cover both interest and acquired

background of knowledge attained from experience or environment). Ideally a list of ambiguous words should be selected whose difficulties are constant in the total sense of practical and non-practical meanings: in this way, the influence of ability and attainment might be eliminated. The assignment of practical and non-practical meanings by the investigator is in some degree arbitrary; but similar arbitrary selections occur in choosing the units of almost every psychological or physical measuring device.

Thirty ambiguous words were selected, and ten ' jokers ' interspersed to obscure the purpose of the test. The series was arranged to avoid any practical connections between neighbouring words. In this preliminary form the test was then given as a ' Meaning of Words Test ' to some 500 boys (aged 10·0 to 11·0) and to twelve candidates for admission to a public day school. The answers were scored P or A according to the *correct* practical or academic meaning given. Very few double responses were given: out of some 12,000 to 15,000 responses (roughly 30 per boy) less than 100 indicated that both a practical and an academic meaning had been given to the same word. 130 of the scripts were selected at random and P and A totals computed for each item. The total (P + A) scores were roughly the same as those for the final version of the test. No item was retained where any difficulty had been found in assigning a clear-cut P or A meaning to the answers. In all seventeen items were retained, and four others added. These were mixed with nine ' jokers '. The final test was as follows: *nut*, *detain*, *port*, *ambush*, *cell*, *portrait*, *tender*, *key*, *spring*, *rule*, *signal*, *flawless*, *switch*, *pick*, *plane*, *joint*, *virtue*, *bulb*, *shaft*, *brace*, *robust*, *fuse*, *strut*, *myth*, *spar*, *cylinder*, *sovereign*, *connection*, *vice*, *humiliation*. It was assumed that, if the items really differentiated between children in the way suggested a good preliminary criterion would be that of sex. They were accordingly administered to boys and girls attending modern schools in an industrial area. Since at the age of 11 + the entrants have had little practical training, a further criterion consisted in contrasting results from children at 11 + and 13 + respectively. There were, at 11 +, 101 boys and 133 girls, and at 13 +, 99 boys and 135 girls, drawn from two boys' and two girls' schools in an industrial area. The P scores were distributed over each group of children in the following ways; the distribution is shown in Table IX. These scores

Table IX. Distribution of P Scores

	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
Girls 11	20	44	33	23	10	1	1	1	—	—	—	—	—	133
Boys 11	3	6	12	19	20	14	8	6	7	3	3	—	—	101
Girls 13	25	26	28	22	16	13	1	3	1	—	—	—	—	135
Boys 13	—	3	2	7	13	16	19	13	10	7	3	4	2	99

are based on different totals (P + A scores) which are distributed at each age level in the manner shown in Table X. The P and (P + A) levels are not independent; and, it was therefore decided to compare the P scores at the (P + A) levels of 11,

14, 17 and 20. Analysing the foregoing data according to the three main components and their interactions, we obtain the following contributions to the total variance (Table XII). The effect of difference in sex is well marked;

Table X. Distribution of (P + A) Scores

		8–	11–	14–	17–	20–
Girls	11	6	31	52	39	5
Boys	11	3	10	19	46	23
Girls	13	3	12	50	55	15
Boys	13	—	4	8	43	44

Table XI. Mean P Scores for Different Sexes at Different Levels

		Ability level			
		20	17	14	11
Girls	11	2·00	2·36	1·79	1·16
Boys	11	6·30	4·20	3·79	2·30
Girls	13	3·13	2·73	1·88	1·50
Boys	13	7·30	5·63	3·63	3·50

Table XII. Analysis of Variance for P Scores

Source of Variation	D.F.	Sums of Squares	Variance	Significance (level exceeded)
Between sexes	1	252506	252506	1%
Between ages	1	18225	18225	5%
Between ability	3	151048	50349	1%
Sex × age	1 × 1	1483	1483	
Age × ability	1 × 3	7158	2386	
Ability × sex	1 × 3	42820	14273	5%
Residual	3	3374	1125	
Total	15	476014		

differences due to ability are the next most significant; those due to age are least significant. Since the interaction between sex and ability is significant at the 5 per cent level, we may infer that the changes in the P score due to sex depend partly on differences in ability, or that the changes resulting from differences in ability depend partly on sex.

The above analysis does not take into account the differing frequencies in the columns of Table X. Accordingly, the P scores obtained by boys and girls of age 13 at the ability level of 20 (in terms of P + A score) have been compared by a method which will do so. Table XIII gives the distributions at this age

Table XIII. Distributions of P Scores

	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
Girls	2	—	4	2	4	2	—	1	—	—	—	—	—	15
Boys	—	—	—	1	2	4	11	7	9	3	3	2	2	44

and level. On comparing the variance between and within sexes, we obtain a variance ratio (F) of 43·4 with 1 and 57 degrees of freedom: the 1 per cent point is only 7·12.

The validity of the homonyms test is less stable than that of the information and vocabulary tests. An indirect indication of the validity is provided by Lemon [9] who compared the homonyms test with the Lambert–Peel information and vocabulary tests. She found a correlation of 0·446 between academic responses in the former and the academic side of the information and vocabulary test, and 0·307 between practical responses and the practical side of the information and vocabulary test. On correlating scores for practical and academic homonyms with a grammar school follow-up in practical and English subjects, she obtained (with 39 pupils taking practical subjects) correlations of 0·147 and −0·148 for P and A scores and (with 83 pupils taking English subjects) correlations of −0·203 and 0·103 respectively). On the other hand, Hart's work (20) does not support the validity of the homonyms test, although it yields a high validity for the information and vocabulary tests.

V. The Relations Between Interest, Ability and Attainment

A verbal test of vocabulary or information can always be regarded as a method of securing data about either attainment, ability, or interest. The purpose of such a test—whether it is meant to test attainment, ability or interest —must be determined by the qualities of the individuals in the sample tested. When such a test is used to assess ability, we assume that the population is homogeneous in interest and educational background. It will fail as an ability test if the children tested are not interested or have suffered from exceptional educational handicaps. Ordinarily these requirements are taken for granted. It is only when we consider the conditions needed for using such a test to measure interest that they are brought to our attention. For such a purpose the population tested should be homogeneous both in ability and in educational background; if not, the form of the test should be so arranged that the requisite conditions are fulfilled as nearly as possible.

When the purpose is to test attainment, we do not as a rule worry about the influence of ability. However, in certain cases, e.g. when we want to investigate causes of backwardness, it is helpful to separate attainment (which may depend largely on past learning and teaching) from the effects due to differences in innate ability. Ability is assumed to be fundamentally innate; attainment to be largely learned; interest to contain an element of motivation. Interest can thus be viewed in two ways: first, it is at once the cause and the effect of activity in any field; our method of using vocabulary and information tests assumes both: secondly, it involves an emotional aspect, which is utilized in tests depending on interest blanks, questionnaires, and rating scales. Further, I assume that attainment is compounded of ability and interest in a multiplicative way. Thus we may express an estimate of attainment by the following equation: $\text{Attainment} = k \times (\text{Ability})^l \times (\text{Interest})^m$, where k, l and m are constants. This

formula differs from that used combining intelligence tests and factors of intelligence: in that case it is generally assumed that the estimate of a criterion score s is made up of a weighted sum of the separate measures, i.e. that $s = w_a a + w_b b + \ldots$, where $a + b$ are scores, and w_a and w_b are constants serving as weights. Broadly speaking, when we combine *different* modes of psychological processes, we should do so multiplicatively; and when (as in factor analysis, prediction of aptitudes, analysis of temperament, and the like) we combine different aspects of the *same*, we do so additively. Possibly we might also combine the latter multiplicatively as well, as suggested by Bartlett [21]; but with this we are not concerned.

The empirical justification for the multiplicative hypothesis rests on several points. First, consider the following cases: if a pupil shows no interest, no attainment follows; if he shows no ability, no attainment is possible; this result follows only with a multiplicative, not with an additive formula. There seems to be something logically and epistemologically incorrect in adding differents; they should be multiplied.

Secondly, I have long been uneasy about the introduction of variables expressing different modes of mental life into the same factor analysis of behaviour. This occurs when workers include such characteristics as persistence, perseverance, interest, and attitude in a matrix where the other constituents are measures of ability or attainment. I doubt even whether it is good scientific method to include ability and attainment in the same matrix: however, this finer criticism lapses if we suppose that ability and attainment are closely interwoven, when interest is held constant, as indeed is tacitly assumed in most factorial studies of human ability and school achievement.

A third argument comes from Hull's analysis of human behaviour [22]. He was concerned with the behavioural components of reaction potential ${}_{S}E_{R}$ (i.e. the tendency for the creature to react in a present learning situation). The components entering into such a situation include the strength of the established habit to react in similar situations (${}_{S}H_{R}$) learned up to the present time, the internal drive (D) induced by making the animal hungry, or the like, the stimulus intensity (V), i.e. the stimulating properties of the learning situation, the amount of incentive reinforcing the response (K), and finally the delay in its enforcement (J). Hull ([22], p. 59) combines these elements multiplicatively in his equation (19): ${}_{S}E_{R} = D \times V \times K \times J \times {}_{S}H_{R}$. These five independent variables are not directly measurable, but are expressed respectively as various exponential and logarithmic functions of the measurable variables —viz. hours of starvation (h), intensity of stimulus (s), weight of food (w), time of delay of reinforcement (t), number of reinforcement (n).

Nevertheless, we should note, first of all, that Hull's formula for the behaviour of a rat includes nothing analogous to the ability term appearing in the expression for a pupil's attainment. Among rats such differences are of academic interest only. In his XVIIIth principle ([22], p. 117), however, Hull recognizes individual differences among his animals, though not as a further variable to

be included in his equation. He deals with the problem by noting that the 'constants' in the equation will carry different numerical values from individual to individual.

With human learners differences in ability become more important, and differences in motivation and drive less important. In the classroom, the incentive term K no longer operates; stimulus intensity (V) is less influential; and J appears only when teachers are slack and unpunctual in returning corrected exercises. On the other hand, drives—in particular secondary drives relating to curiosity, anxiety, and dominance—exercise a more powerful influence. The combined action of all those elements which thus approximately resemble Hull's K, V, J and D might well be designated 'interest'. Accordingly, if we assume that Interest $= k(D \times K \times V \times J)$, only one term remains unexplained, viz. ${}_{S}H_{R}$. This measures the established strength of response at the beginning of the experiment. In the classroom it would be measured by the working attainment of the pupils up to the point at which they start the new learning. Thus, proceeding on the analogy of Hull's equation, and incorporating a sixth term to represent individual differences, we arrive at the following expression for the potential attainment of a pupil in a new learning situation:

$$\text{Attainment} = k(\text{Attainment-to-date} \times \text{Interest} \times \text{Ability}).$$

Measures of the three terms on the right-hand side of the equation can be obtained by suitable tests, examinations, or rating scales. Analysis along these lines is being attempted; but it is too early as yet to discuss the results.

References

[1] Peel, E. A. (1948). Assessment of interest in practical topics. *Brit. J. Educ. Psychol.*, XVII, 41–47.

[2] Lambert, C. M. (1949). Symposium on selection of pupils for different types of secondary school: VII. A survey of ability and interest at the stage of transfer. *Brit. J. Educ. Psychol.*, XIX, 67.

[3] Watts, A. F., and Slater, P. (1950). The allocation of primary school leavers to courses of secondary education. *National Foundation of Educational Research*, Report 2.

[4] Angus, L. (1949). *A Comparative Study of the Methods of Interest Measurement in Science and its Relation to Ability and Achievement.* M.Ed. Thesis, Durham.

[5] Pennington, A. (1950). *A Study of Leisure Time Interests.* M.Ed. Thesis, Durham.

[6] Pennington, A. (1950). A study of leisure interests in schoolboys. *Research Review*, I, 31–37.

[7] Cureton, L. A. (1955). *A Follow-up Study of the Lambert–Peel Practical Interest Tests in Certain Secondary Modern Schools of the West Riding.* Dissertation for Diploma in the Psychology of Childhood, Birmingham University.

[8] Potgieter, M. (1954). *The Determination of the Predictive Value of Academic and Practical Interest Scores for Allied Attainment in the Grammar School.* Dissertation for Diploma in the Psychology of Childhood, Birmingham University.

[9] Lemon, G. (1955). *Information Gained from a Population of Secondary Modern and Grammar School Children in the West Riding of Yorkshire.* Dissertation for Diploma in the Psychology of Childhood, Birmingham University.

[10] Burroughs, G. E. R. (1955). *Student Teachers' Interests.* Unpublished research, Birmingham University.
[11] Miller, T. W. G. (1958). *A Critical and Empirical Study of the Emergence Development and Significance of the Comprehensive Secondary School in England.* Ph.D. Thesis, Birmingham University.
[12] Fryer, D. (1931). *The Measurement of Interests in Relation to Human Adjustment.* New York: Holt.
[13] Slater, P. (1944). *Selective Vocabulary Test.* London: Harrap.
[14] Wyman, J. B. (1925). Tests of intelligence, social and activity interests. Ap. L. M. Terman, *Genetic Studies of Genius.* Stanford University.
[15] Watts, A. F., and Slater, P. (1950). The allocation of primary school leavers to courses of secondary education. *National Foundation for Educational Research,* Report No. 11, Newnes, London.
[16] Evans, G. (1958). Secondary school pupils' interests in arithmetic. Dissertation for Diploma, Birmingham University.
[17] Thurstone, L. L. (1948). The Rorschach in psychological science. *J. Abn. Soc. Psych.,* XLIII, 471–475.
[18] McNemar, Q., and Merrill, T. (1942). *Studies in Personality.* New York: McGraw-Hill.
[19] Foley, J. P., and MacMillan, Z. L. (1943). Mediated generalisation and the interpretation of verbal behaviour as related to differences in professional training. *J. Exp. Psychol.* XXXIII, 299–310.
[20] Hart, B. T. (1959). *The Relationship between the Interests of Secondary Modern Boys as measured by Information and Vocabulary Tests and their School, Class and Intelligence.* Dissertation for Diploma in the Psychology of Childhood, Birmingham University.
[21] Bartlett, M. S. (1953). Uppsala Symposium on factor analysis: factor analysis in psychology as a statistician sees it. *Nordisk Psykologi Monograph Series,* No. 3.
[22] Hull, C. L. (1951). *The Behaviour System.* New Haven: Yale University Press.

Vol. XII Part II The British Journal of Statistical Psychology November 1959

AN ANALYSIS OF THE LIMITING METHOD

By J. Brown and V. R. Cane
Birkbeck College, University of London Newnham College, University of Cambridge

The value for a sensory threshold which the Limiting Method yields is shown to be mathematically dependent on the details of the experimental procedure: in particular, it is highly sensitive to the size of the steps between successive values of the variable stimulus. Moreover, it is impossible to ensure in advance that any two thresholds are obtained by equivalent experimental procedures, except in the trivial case where the relevant psychophysical distributions are already known. Nevertheless, on certain assumptions, the method may be validly used to test whether thresholds change in an expected manner; and the way in which this can be done is described. The Method of Average Error and the Békésy–Oldfield method for the continuous recording of threshold are basically similar to the Limiting Method.

I. Introduction

In the Limiting Method a variable stimulus is altered by steps until the subject makes a prescribed judgment, such as ' present ', when an absolute threshold is being measured, or ' equal to the standard ', when a difference threshold is involved. This stopping point is then taken as providing one estimate, T, of the threshold. The final estimate will be the mean value $\bar{T}$ obtained from a series of such runs. $\bar{T}$ has been assumed by most previous writers to have a stable value. In general, however, this assumption is unwarranted. As we propose to demonstrate, the expected value of T, when obtained from a given run, depends on the particular testing points chosen. It follows that both T and $\bar{T}$ will be to some extent functions of the particular procedures adopted for their determination.

When the variable is presented at any particular testing point, x, there is a certain probability, p, that the subject tested will then give the particular reply which ends the run. If, for example, the judgment is ' equal ', this probability will be very low when the difference between the variable and the standard is large and high when the difference is small. Strictly speaking, however, this analysis holds only for the case where the value of p remains unaffected by the testing points that may previously have been used: for other cases a similar analysis could be given provided the manner in which p depends on previous testing points were known. Suppose that the variable is to be presented in succession at x_1, x_2 ,... x_r, ..., until the trial ends. Let p_1, p_2, ... p_r, ... denote the corresponding values of p, and P_1, P_2, ... P_r ... the probabilities that the rth trial will end at each of these values of x. Then clearly $P_1 = p_1$. However,

$P_2 \neq p_2$, unless $p_1 = 0$: for, if p_1 is not zero, the trial may have ended at x_1. Hence $P_2 = p_2(1-p_1)$ or $P_2 = p_2 q_1$. In general,

$$P_r = p_r q_1 q_2 \cdots q_{r-1}, \tag{1}$$

where $q_i = 1-p_i$ for $i = 1, 2, \ldots n$.

Now P_r is the probability that the stopping point has the value x_r. Hence the expected value of the stopping point is

$$T = \sum P_r x_r. \tag{2}$$

If the same values of x are used for all the runs, this will also be the expected value of $\bar{T}$: otherwise, its expected value will be the mean of T when (2) is applied to the individual trials.

From (1) and (2) it is evident that, unless a highly improbable assumption is made about the distribution of p (see below), the expected values of $\bar{T}$ depends on the particular testing points chosen for the various runs. Hence, no two thresholds obtained by the Limiting Method are necessarily comparable. In consequence the conventional procedure and treatment of this method is liable to produce spurious constant error changes in threshold from one sitting to the next, or a breakdown in Weber's Law.

The conventional treatment of the Limiting Method seems to be based on a tacit assumption that, in any single run, p is zero for all values of x up to a critical value at which it suddenly becomes unity. In this case, T would occur at the first testing point after this critical value. Its exact amount would depend on the choice of this testing point, but not on the choice of the earlier points. It is unlikely, however, that the distribution of p is ever a step function, except perhaps for very rapid runs.

II. Urban's Analysis

F. M. Urban (1908) derived the formulae (1) and (2) given above[1]. Faulty reasoning then led him to claim that $\bar{T}$ would tend to occur at the value of x for which $p = \frac{1}{2}$, provided different values of the variable are used in successive trials. If this were correct $\bar{T}$ would, of course, have a stable value, with the proviso stated. The acceptance of Urban's views presumably accounts for the unguarded use of $\bar{T}$ as a valid threshold measure down to the present day.

Urban reasoned as follows. In a symmetrical distribution, the mean value corresponds to the most probable value (the mode). The distribution of T cannot be assumed to be symmetrical for any particular set of values of x. But if different values of x are used on successive trials the mean, $\bar{T}$, of the resultant values of T should have a normal, and hence a symmetrical, distribution. To reach his final conclusion, Urban then made two mistakes. First, he supposed that the mode of T, i.e. the maximum value of P, would always tend to occur where

[1] In *The Essentials of Mental Measurement* (3rd ed., 1925, chapter III) Brown and Thomson give a critical account of Urban's analysis of the traditional psychophysical methods and a summary of Thomson's valuable contributions to the theory of the Limiting Method.

$p = \frac{1}{2}$: that this is incorrect will be demonstrated in a moment. Secondly, he tacitly assumed that the mode of $\bar{T}$ necessarily coincides with the supposed common mode of the contributory distributions: once this assumption is stated explicitly its falsity is evident. It is easy to show that, while the maximum value of P cannot occur later than the first testing point for which $p \geqslant \frac{1}{2}$, it may in fact occur much earlier. We have

$$\frac{P_{r+1}}{P_r} = \frac{p_{r+1} q_1 \cdots q_{r-1} q_r}{p_r q_1 \cdots q_{r-1}} = \frac{p_{r+1} q_r}{p_r}, \qquad \text{by (1).}$$

Now let

$$p_r = \frac{1}{u}, \qquad p_{r+1} = \frac{1}{u-\delta}$$

where $u \geqslant 1$ and $0 \leqslant \delta \leqslant (u-1)$;

then

$$\frac{P_{r+1}}{P_r} = \frac{1}{u-\delta} \frac{u-1}{u} \frac{u}{1} = \frac{u-1}{u-\delta} \tag{3}$$

Hence

(i) if $\delta > 1$, $P_{r+1} > P_r$,

(ii) if $\delta < 1$, $P_{r+1} < P_r$,

(iii) if $\delta = 1$, $P_{r+1} = P_r$.

If we know how p varies with the testing point, relation (ii) shows how to choose successive testing points so that P_1 will be the maximum value of P; and relation (i) shows how to choose them so that the maximum value is as late as possible. But, since $\delta \leqslant u - 1$, $\delta \leqslant 1$ if $u \leqslant 2$, i.e. if $p \geqslant \frac{1}{2}$: in other words, the maximum value of P cannot occur later than the first value of x for which $p \geqslant \frac{1}{2}$, as stated before. Relation (iii) shows how to choose successive testing points so that P is constant so long as $u \geqslant 2$. It is therefore quite wrong to suppose that the maximum value of P necessarily tends to occur at that value of x for which $p = \frac{1}{2}$.

Are the likely variations in $\bar{T}$ with procedure sufficient to matter? And if they are, how is it possible to use the Limiting Method effectively? The answers to these questions will be considered in the next two sections.

III. The Effects of Different Procedures on the Mean Value

It will be sufficient to consider the effect of the testing points on the expected value of T, since the expected value of $\bar{T}$ is dependent on those of T. We consider first the case in which the steps between the successive values of x have a constant size. This means that the testing points will be completely determined by (i) the size of the steps and (ii) the starting point.

(i) If we assume that the distribution of p is cumulative normal, with standard deviation σ, and that the starting point is at -3σ, then the table shows how the expected value of T varies with the step size. It will be seen that the step size has a considerable effect on the value. As the step size is reduced, the expected value of T regresses towards the starting point. It can be shown that if n is the

number of steps between -3σ and $+3\sigma$, then the expected value of T decreases as $1/\sqrt{n}$ (Cane 1956). For n between 4 and 25, the expected value is approximately $(4{\cdot}86n^{-1/2}-1{\cdot}67)\sigma$.

THE EFFECT OF STEP SIZE (STARTING POINT AT -3σ)

Step size	3σ	1σ	$0{\cdot}5\sigma$	$0{\cdot}25\sigma$
Expected value of T	$1{\cdot}5\sigma$	$0{\cdot}27\sigma$	$-0{\cdot}24\sigma$	$-0{\cdot}67\sigma$

(Expressed as a deviation from the centre of the distribution of p.)

(ii) The starting point is usually chosen at a value of x for which p is negligibly small. In this case its main importance lies in the fact that it will determine the values of later testing points in the range when p is not negligible. Hence two procedures, each adopting the same step size and a starting point at which p is negligible, will differ most when the testing points involved are half a step out of phase. However, even then, the expected values of T obtained with the two procedures will usually be nearly the same. For example, if the step size is 1σ and the distribution of p normal, then the expected value of T will be $0{\cdot}2728\sigma$ when the starting point is -3σ and $0{\cdot}2763\sigma$ when it is $-2{\cdot}5\sigma$. When the value of p is not negligible at the starting point, the latter naturally has a greater effect. For example, if the step size is $0{\cdot}5\sigma$ and the distribution of p normal, the expected value of T will be $-0{\cdot}2444\sigma$ when the starting point is at -3σ, and $-0{\cdot}19\sigma$ when it is at $-1{\cdot}5\sigma$. There are two situations in which p is unlikely to be negligible at the starting point: (i) in the case where ascent or descent from the variable equal to the standard is used and (ii) in the method introduced by Békésy (1947) and independently by Oldfield (1949) for the continuous recording of threshold change. The conclusion is that two procedures are acceptably equivalent if the step size in σ units is the same for both, and if the values of p at both starting points are negligibly small. But this is only a sufficient condition, not a necessary condition, for acceptable equivalence.

So far we have been considering cases where the variable is altered in steps of constant size. If the size varies, the mean step size will have roughly the same effect as a constant step size of the same magnitude. However, this applies only if variation in the step size is randomly distributed throughout the trial: it would not apply, for example, if the steps were initially large and then became small when the experimenter expected the stopping point to be reached. Indeed, if the experimenter chooses when to start using very small steps, he can effectively determine when the stopping point occurs.

It may be wondered how the Limiting Method has been able to yield reasonably satisfactory results, if the foregoing analysis is correct. The answer probably is that experimenters have tended to use their own perception in order to decide what is a reasonable step size under the conditions which they wished to compare: this has tended to make the step size in σ units roughly constant, so that roughly comparable threshold measures have been obtained. Such a procedure is clearly risky and unreliable.

IV. A Valid way of using the Method

If $\bar{T}$ cannot be assumed to have a stable value, one way out of the difficulty is to stop using $\bar{T}$ as a threshold measure, and to use the results to estimate the distribution of p instead. If the same testing points are retained for successive trials, the results can then be treated in almost the same way as results obtained by the constant method. The only complication is that the variable will have been presented more often at earlier testing points than at later ones. The disadvantage of this solution is that the main attraction of the Limiting Method, namely, the small amount of calculation required to obtain the threshold measure, $\bar{T}$, is then lost. It is therefore worth while to seek another solution to the problem.

In devising a valid way of using the method, two problems have to be solved. The first is that $\bar{T}$ is expected to occur at the same value of p under two different conditions only if the same step size (in σ units) is used under both conditions. Now, provided a definite hypothesis is to be tested as is usually the case, one solution is to alter the step size in accordance with the hypothesis. For example, if the standard deviation is expected to be twice as large under condition 1 as under condition 2, then the step size is made twice as large under condition 1 as under condition 2. It will be shown that the results expected are obtained only if the hypothesis is true, and that deviant results will show the direction of any breakdown of the hypothesis. The sole assumption necessary is that the shape of the distribution of p remains constant no matter what changes its standard deviation, σ, may undergo. This assumption is likely to hold to a sufficient degree of approximation, except perhaps when thresholds close to the absolute thresholds are being measured. The second problem is this. To use $\bar{T}$ as a threshold measure it is necessary to know where the centre point of the distribution of p lies[1]. But in practice we do *not* know the whereabouts of this point. The best solution to this problem seems to be to determine *two* mean transition points, one $\bar{T}_a$ for ascent of the variable to the standard stimulus and one $\bar{T}_d$ for descent. The interval $(\bar{T}_d - \bar{T}_a)$ can then be used as a threshold measure, provided $\bar{T}_d$ and $\bar{T}_a$ are complementary.

The proposed method is as follows: (for convenience of exposition sampling errors will be ignored until later). Under each condition to be compared—for example, for each value of the standard—the mean transition points $\bar{T}_a$ and $\bar{T}_d$ for ascent and descent are determined. If ascents are terminated when the subject passes from the judgment J to not-J, then descents will be terminated when he passes from not-J to J. Thus $\bar{T}_a$ and $\bar{T}_d$ will be strictly complementary. Let p denote the probability of J at any given value of the variable so that the corresponding value of not-J will be $(1-p)$. The same step size (or sizes, see

[1] In the traditional procedure the distance of $\bar{T}$ from the value of the standard stimulus is taken as the threshold measure. This distance divides into the distance d_1 of $\bar{T}$ from the centre point of the distribution of p and the distance d_2 of this point from the standard. It should be noted that d_2 is not necessarily a function of the standard deviation, σ, and also that d_2 is unlikely to be zero except for a transition from a judgment of variable less to variable greater or vice versa.

later) is used for ascents and descents. Further, ascents are always begun where p is very small, and descents where $(1-p)$ is very small. The interval $(\bar{T}_d - \bar{T}_a)$, which is the threshold measure, will be denoted I. The general problem is to test whether $\sigma_1 = k\sigma_2$, where σ_1 and σ_2 are the standard deviations of the distribution of p under the two conditions c_1 and c_2 respectively. This is achieved by using a step size under condition c_1 which is k times the step size under condition c_2. If then $\sigma_1 = k\sigma_2$, the threshold interval I_1 under condition c_1 should be k times as large as the threshold interval I_2 under condition c_2: for, when $\sigma_1 = k\sigma_2$, given values of p and $(1-p)$ will be k times as widely separated under c_1 as under c_2. If $\sigma_1 > k\sigma_2$, then I_1 will be greater than kI_2: for now given values of p and $(1-p)$ will be more than k times as widely separated under c_1 as under c_2; moreover, in terms of σ units, a smaller step size will have been used under c_1 than under c_2, and the effect of this is to make $\bar{T}_a$ and $\bar{T}_d$ occur at smaller, and hence more widely separated values of p and $(1-p)$, as the table shows. Similarly, if $\sigma_1 < k\sigma_2$, then I_1 will be less than kI_2. Thus the method makes it possible to test any hypothesis concerning the relation between σ_1 and σ_2, and to see in what direction, if any, the hypothesis breaks down. It does not enable the magnitude of any breakdown to be assessed accurately, since any breakdown alters the effective step size in such a way as to increase the change in the threshold interval I.

A numerical example may help to make this clear. Let us suppose we might wish to test whether $\sigma_1 = \sigma_2$, and that the step size chosen is $0{\cdot}5\sigma_1$, and that $\sigma_2 = 2\sigma_1$ (contrary to the hypothesis to be tested). Then the step size will be $0{\cdot}5\sigma_1$ under condition c_1 and $0{\cdot}25\sigma_2$ under condition c_2. If p is normally distributed, then, as the table shows I_1 will be $2 \times 0{\cdot}24 \times \sigma_1$ and I_2 will be $2 \times 0{\cdot}67 \times \sigma_2$. Since $\sigma_2 = 2\sigma_1$, I_2 will thus be between 5 and 6 times as large as I_1. It is important to pay attention to the sign of I_1 and of I_2. If in this example we had had $\sigma_2 = 0{\cdot}5\sigma_1$, then the step size under condition c_2 would have been σ_2; and I_2 would have been negative, viz. $-2 \times 0{\cdot}27\sigma_2$, as can be seen from the table.

We have described the method as though it were essential to use only a single step size. In order to use several step sizes it is merely necessary to ensure that the procedure outlined it represented for each step size. The use of several step sizes has psychological advantages and may also be convenient for the following reason. Suppose we wish to test whether $\sigma_1 = \sigma_2$, and find that $I_1 > I_2$. This tells us that σ_1 is larger than σ_2, but not how much larger. If several step sizes are used, examination of individual values of I_1 and I_2 may indicate the actual relation between σ_1 and σ_2. For example, if we find $I_1 = 2I_2$ when the step size is twice as large under condition c_1 as under condition c_2, this indicates that $\sigma_1 = 2\sigma_2$.

In practice sampling errors must necessarily affect the observed values of I. The problem is then to determine whether an observed relation between I_1 and I_2 differs significantly from an expected relation. This can be solved either by an analysis of variance, or, if a variety of step sizes are used, by an application of a t-test to the differences between corresponding values of I_1 and I_2.

Two further points should be noted. First, if the distribution of p is antisymmetrical (i.e. symmetrical in the manner of a linear or a cumulative normal distribution), then $\bar{T}_d$ and $\bar{T}_a$ will lie at equal distances from the centre point of the distribution, provided equal step sizes have been used. The mean of $\bar{T}_d$ and $\bar{T}_a$ therefore tells us where this centre point lies. Secondly, $\bar{T}_a$ and $\bar{T}_d$ do *not* correspond to the upper and lower thresholds obtained by the traditional application of the Limiting Method. $\bar{T}_a$ is the mean transition point for passing from J to not-J, and $\bar{T}_d$ for passing from not-J to J. Hence $\bar{T}_a$ cannot be interpreted as relating to the transition from 'less than' to 'equal' nor $\bar{T}_d$ as relating to the transition from 'greater than' to 'equal'. Indeed, upper and lower values of $(\bar{T}_d - \bar{T}_a)$ can be obtained by defining J either as 'less than' or as 'greater than' respectively.

V. Further Criticisms

It has been alleged that the Limiting Method is seriously affected by an error of expectation. This error is often said to consist in the fact that the subject reports that the variable and the standard are equal before objective equality is reached. However, even if the distribution of p is centred at the point of equality, this point is only the expected value of $\bar{T}$ when a particular step size is used: moreover, in a normal distribution it is rather a large step size (see table). Further, since the stopping point is normally at the transition from a judgment of 'not equal' to one of 'equal', the distribution of p is, in practice, likely to be centred somewhere on the approach side to the standard and not at the point of equality to the standard. To show that there is a real error of expectation it is necessary to prove that the values of p at later testing points are affected by presenting the variable at earlier testing points. It should be noted that an error of expectation will not necessarily undermine the validity of the new way of using the method described in the previous section. If the error is constant for a given step size (in σ units), we may still expect $I_1 = kI_2$, provided that $\sigma_1 = k\sigma_2$, and that the absolute step size is k times as large when determining I_1 as when determining I_2.

In conclusion, let us consider Oldfield's version of the method. This was introduced as a way of recording fluctuations in the threshold from moment to moment. A variable stimulus is changed at a constant rate until the subject perceives, or ceases to perceive, a certain feature of the stimulus; he then presses a button which reverses the direction of change, and the complementary half of the cycle ensues. This resembles the Limiting Method in that the variable changes independently of the subject and in a constant direction until he makes a certain reaction (pressing the button). We denote the end-point of an ascent by T_a and of a descent by T_d. An important feature is that, since ascents will usually end before $p = 1$, descents will start where $(1 - p)$ is appreciable; similarly, ascents will start where p is appreciable. This will have two consequences. First, successive values of T_a and T_d will tend to be correlated. Secondly, because the starting points vary, no two values of T_a or T_d will necessarily be comparable. The following model indicates the results likely to be obtained

with the method. The subject is assumed to judge the variable, not continuously, but at equally spaced points $x_1, x_2, \ldots x_n$; successive judgments are assumed to be independent; finally, the distribution of p is assumed to be linear from $p = 0$ at x_1 to $p = 1$ at x_n. Since $p = 0$ at x_1, T_a cannot occur at x_1; and, since $(1-p) = 0$ at x_n, T_d cannot occur at x_n. For positions $x_2 \ldots x_n$ with ascents and $x_1 \ldots x_{n-1}$ with descents, the probabilities of T_a and T_d respectively follow the binomial distribution defined by probability $\frac{1}{2}$ and $N = n-2$. The theoretical correlation between successive stopping points can be shown to be

$$1 - \frac{(n-3)!}{2^{n-4}(n-1)^{n-2}} \sum_{k=0}^{k=n-3} \frac{(2n-2)^k}{k!}.$$

The correlation is about two-thirds for $n = 11$.

The Method of Average Error, in which the subject himself adjusts the variable stimulus, is not unlike the Limiting Method, especially in the Békésy–Oldfield form. As with the conventional procedure, the threshold obtained in any trial will depend on the starting point and the step size. However, since the step size is here controlled by the subject, it is difficult to see how the method could be modified so as to yield a valid threshold measure. There is the further complication that the subject may not make his adjustments in a constant direction.

VI. Conclusion

Despite its popularity, the reliability of the Limiting Method as ordinarily used is much inferior to that of the Constant Method. The modified version which we have suggested should eliminate the fundamental source of this unreliability, and is almost as easy to carry out as the traditional procedure. Compared with the Constant Method, the Limiting Method has the advantage that sharp contrasts between successive values of the variable are avoided; this diminishes the risk that the results may be affected by sensory adaptation or by habits in guessing. Accordingly, in the modified form the Limiting Method may well prove to be the most reliable of all the psychophysical procedures.

References

[1] Cane, V. R. (1956). Some statistical problems in experimental psychology. *J. Roy. Statist. Soc.*, XVIII, 177–201.

[2] Békésy, G. V. (1947). A new audiometer. *Acta. Oto-Laryng. Stokh.*, XXXV, 411–422.

[3] Oldfield, R. C. (1949). Continuous recording or sensory thresholds and other psychophysical variables. *Nature*, CLXIV, 581.

[4] Urban, F. M. (1908). *The Application of Statistical Methods to the Problems of Psychophysics*. Philadelphia: Psychological Clinic Press.

Vol. XII Part II The British Journal of Statistical Psychology November 1959

THE ATHLETIC ABILITY OF BOYS IN A SECONDARY MODERN SCHOOL

By G. Highmore and B. W. Jones
Sawston Village College Loughborough Training College

A group consisting of just over a hundred boys were tested and retested at the ages of eleven and thirteen respectively with eight tests of athletic ability. An analysis of the correlations by means of Burt's procedure furnished a factor matrix consisting of a basic factor for general athletic ability and three group factors, slightly overlapping, for running, jumping, and throwing respectively. Contrary to what is usually found in factorial studies with cognitive tests, the contribution of the general factor to the total variance increased appreciably during the two-year interval between the two investigations : at the same time the group factors became more specific and more distinctly recognizable. Three tests, one drawn from each of the three group factors, proved sufficient to yield a broad indication of a pupil's athletic ability during the secondary school stage.

I. Problem

The Differentiation of Abilities. Several investigations have been made into the effects of age on the differentiation of mental abilities (Burt, 1954 and ref.) ; but so far no comparable inquiry appears to have been carried out in the case of physical abilities. The experiments described in the following paper were planned as a preliminary study of this aspect of the general problem.

The present investigation is a sequel to the factor analysis of athletic ability described by Highmore and Taylor in an earlier article in this *Journal* (VII, 1954, pp. 1–8). The two studies, taken together, form part of a comparative inquiry undertaken to determine the effect of increasing age on the development of the special abilities underlying the psycho-physical processes involved in athletics and games. In the earlier investigation 110 boys, aged 11 to 12, were tested; and the same group was still available, with the same teachers for physical education, when the second investigation was started. Of the entire group 102 completed the whole series of tests ; but to simplify computation the number was reduced to 100.

On the first occasion the boys were tested during their first athletic season in a secondary modern school ; and as long a period as possible was allowed to elapse before the tests were repeated. Since many of them reached their fifteenth birthday early in their final year and were expected to leave in December, it was necessary to carry out the second set of tests during the preceding athletic session. Thus the total interval amounted to two athletic seasons ; and, when tested for the second time, the boys had reached the age of 13 to 14 years.

II. Methods

Tests and Criteria. The following tests were used on both occasions : (1) a 50-yard sprint on a grass playing field ; (2) a run of 150 yards made on a gently curving track ; (3) a standing long jump (three jumps were allowed, and the best recorded) ; (4) a high jump (three trials were allowed at each height, and success was recorded if the boy succeeded in one trial out of three) ; (5) a standing upward jump (three jumps were allowed, and the best recorded) ; (6) football throw-in (three throws were made, and the best recorded) ; (7) throwing the cricket ball (a $5\frac{1}{2}$ oz. ball was used, and a preparatory run allowed) ; (8) football punt for distance (a dead association football, No. 4 size, was kicked from a marked line, and the best of three kicks accepted).

Before the tests were carried out, the senior physical education master and his assistant were asked to place the boys in order of merit according to the all-round performance of each in the traditional branches of their subject. Their ratings formed the criterion.

Correlations and Factor Analysis. As in the earlier investigation Highmore supervised the testing. The raw scores were then passed to Mr. Knott and Mr. Sussex, who were good enough to undertake the mechanical work of calculating the coefficients of correlation and other constants. The correlations are shown in Table I : only three coefficients do not exceed twice their standard error.

Table I. Correlations Between Tests

Test	1	2	3	4	5	6	7	8	9
1. Sprinting (50 yards)	—	0·68	0·49	0·49	0·31	0·52	0·47	0·60	0·52
2. Running (150 yards)	0·68	—	0·54	0·52	0·33	0·54	0·43	0·39	0·45
3. Long Jump	0·49	0·54	—	0·59	0·41	0·61	0·46	0·47	0·39
4. High Jump	0·49	0·52	0·59	—	0·33	0·64	0·49	0·42	0·54
5. Standing Upward Jump	0·31	0·33	0·41	0·33	—	0·16	0·17	0·14	0·20
6. Throwing Football	0·52	0·54	0·61	0·64	0·16	—	0·71	0·61	0·62
7. Throwing Cricket Ball	0·47	0·43	0·46	0·49	0·17	0·71	—	0·54	0·61
8. Punting Football	0·60	0·39	0·47	0·42	0·14	0·61	0·54	—	0·43
9. Criterion	0·53	0·45	0·39	0·54	0·20	0·62	0·65	0·43	—

The standard errors vary from $\pm 0{\cdot}05$ for a correlation of 0·68 to $\pm 0{\cdot}10$ for a correlation of 0·17.

The factorization of the correlation table was carried out by the same three methods as were used in the preceding investigation. (i) A preliminary analysis was first made by the ordinary method of simple summation. Burt's formula (1917, p. 53) was used[1]. The communalities were obtained by successive approximation, and for this purpose six iterations were needed. One

[1] This is the procedure commonly used by American investigators who have applied factorial techniques to the study of athletic abilities. They generally describe it as the ' centroid method ' (the name proposed by Thurstone) ; but Thurstone's formula is identical with Burt's. It is chiefly in the method of rotating the summational or ' centroid ' factors that the two writers differ.

TABLE II. FACTOR SATURATIONS. FIRST INVESTIGATION

Type of factor method of computation	Bipolar factors Simple summation			Non-overlapping group factors Direct group factor analysis				Overlapping group factors Arithmetical rotation			
Sprinting (50 yards)	0·734	0·281	0·068	0·638	0·237	—	—	0·595	0·506	0·027	0·128
Long Jump	0·665	0·222	0·176	0·601	0·237	—	—	0·614	0·438	0·039	−0·084
Running (150 yards)	0·562	−0·369	−0·123	0·412	—	0·581	—	0·354	0·491	0·014	0·248
High Jump	0·634	0·031	−0·066	0·665	—	0·222	—	0·576	0·152	0·023	0·343
Standing Upward Jump	0·543	0·320	−0·055	0·409	—	0·419	—	0·394	0·425	0·018	0·252
Throwing Cricket Ball	0·644	−0·339	0·164	0·587	—	—	0·426	0·602	0·032	0·452	0·063
Punting Football	0·612	−0·228	0·118	0·609	—	—	0·285	0·621	0·013	0·228	0·096
Throwing Football	0·601	−0·401	−0·086	0·517	—	—	0·540	0·494	−0·028	0·554	0·181
Criterion	0·759	−0·255	−0·196	0·764	—	—	0·435	0·625	0·110	0·319	0·346
Square Sum	3·732	7·61	1·37	3·114	0·112	0·563	0·734	2·716	9·07	7·93	4·30
Contribution to Variance (per cent)	41·5	8·5	1·5	37·7	1·4	6·8	8·9	30·2	10·1	8·8	4·8

TABLE III. FACTOR SATURATIONS. SECOND INVESTIGATION

Type of factor method of computation	Bipolar factors Simple summation			Non-overlapping group factors Direct group factor analysis				Overlapping group factors Arithmetical rotation (First approximation)			
Sprinting (50 yards)	0·780	−0·093	0·406	0·773	0·346	—	—	0·798	0·344	−0·041	−0·145
Running (150 yards)	0·719	−0·270	0·172	0·725	0·346	—	—	0·731	0·259	−0·072	0·110
Long Jump	0·743	−0·209	−0·247	0·718	—	0·318	—	0·692	0·007	0·073	0·415
High Jump	0·738	−0·056	−0·027	0·752	—	0·160	—	0·676	−0·018	0·190	0·313
Standing Upward Jump	0·379	−0·401	−0·082	0·318	—	0·568	—	0·391	0·118	−0·216	0·312
Throwing Football	0·836	0·302	−0·165	0·752	—	—	0·459	0·734	−0·092	0·505	0·130
Throwing Cricket Ball	0·732	0·403	0·001	0·615	—	—	0·619	0·646	−0·041	0·524	−0·063
Punting Football	0·649	0·150	0·086	0·615	—	—	0·256	0·607	0·075	0·275	−0·028
Criterion	0·682	0·179	0·046	0·642	—	—	0·292	0·630	0·047	0·317	−0·004
Square Sum	4·48	0·598	0·342	4·039	0·239	0·449	0·746	3·98	0·218	0·801	0·422
Contribution to Variance (per cent)	49·8	6·6	3·8	44·9	2·7	5·0	8·3	44·2	2·4	8·9	4·7

general and two bipolar factors were extracted. The residuals then remaining proved to be non-significant. (ii) Following Burt's procedure the lines of division indicated by the bipolar factors were used to determine the grouping of the tests for the group factors ; and three non-overlapping group factors, with one basic factor, were then computed by the ' group factor method '. (iii) On the basis of these two analyses an arithmetical rotation matrix was computed and used to obtain a matrix of overlapping factors. The entire procedure was thus identical with that already described in this *Journal*, which in turn was adapted from that used by Burt and his collaborators in their earlier studies of athletic abilities (Highmore and Taylor, 1954, p. 758). The correlation table reconstructed from the results of this final analysis gave an excellent fit to the table of observed coefficients.

In order that the factor analysis should be carried through quite objectively, these further calculations were undertaken by the second author (B. W. Jones) who remained ignorant of the identity of the tests or of the conclusions reached in the earlier research until he himself had completed the computations. The results of the earlier research are reproduced for comparison in Table II ; and those of the present investigation are set out in Table III.

III. Conclusions

A comparison of Tables II and III shows that the total common factor variance has appreciably increased. Much of the increase appears to be due to the common group factors.

The classification of tests suggested by this second investigation is somewhat similar to that inferred from the results obtained from the same children at the younger age. On the whole the previous results, as pointed out by Highmore and Taylor, appeared to confirm the conclusions reached in the earlier researches of Burt and his collaborators. They too found distinct group factors for running, jumping and throwing. For these various activities, not only are different muscles involved, but different forms of muscular coordination. As Burt observes, throwing depends chiefly on arm muscles, and, when aiming is involved, demands dexterity ; running depends chiefly on leg muscles, and demands muscular speed and length of stride ; jumping depends chiefly on the strength of the leg muscles, but length of leg also facilitates such movements, though in a minor degree. It is therefore possible that the increase in the group factor variance may in part be due to the maturation of physical processes underlying athletic performances—a maturation which, according to some, is much accelerated by the approach of adolescence ; but it seems far more probable that the results observed are chiefly due to continued practice.

The contribution of the large ' general ' or ' basic ' factor also appears to have increased with increase in age. This at first sight would seem to contradict Burt's theory of progressive differentiation ; but, as he points out, athletic ability is itself to be regarded as a special rather than a general ability, and its increase in variance is therefore itself a manifestation of increasing differentiation.

Finally, if we wish to make an estimate of a boy's general athletic ability, it would seem possible to do so from three tests only. In the earlier investigation the three tests suggested by Highmore and Taylor were the 50-yard sprint, the football throw-in and the high jump, giving equal weight to each test. The results of the present study, however, suggest that a better selection would be the 50-yard sprint (which is easily administered and is the most representative of the factor for running events), the long jump rather than the high jump, and the cricket-ball throw.

References

Burt, C. (1917). *The Distribution and Relations of Educational Abilities.* London : P. S. King.

Highmore, G., and Taylor, W. R. (1954). A factor analysis of athletic ability. *Brit. J. Statist. Psychol.* VII, 1–8.

Burt, C. (1954). The differentiation of intellectual ability. *Brit. J. Educ. Psychol.*, XXIV, 76–90.

Vol. XII Part II The British Journal of Statistical Psychology November 1959

TWINS IN THE SCHOOL POPULATION

By F. Sandon
Chief Examiner, Somerset Education Committee
Formerly Selection Officer, Birmingham Education Committee

The following paper is based on a statistical analysis of marks gained by over a thousand twins who have sat for various 11+ examinations for transfer to grammar schools. It is shown that the average ability of twins is well below that of the general population. The correlations between twins of the same sex are much higher than those between twins of opposite sex.

I. Sources of Data

In recent discussions, in this *Journal* and elsewhere, on the inheritance of mental abilities [4, 5], the data obtained from twins have formed an important part of the evidence. They have, however, been subjected to considerable criticism from educationists and others ; and the conclusions drawn have at times been attributed to the preconceived opinions of those who have carried out the researches. The figures reported here have been gathered in the course of routine work for two different education authorities, and may therefore be regarded as entirely free from any suspicion that the planning of the inquiry may have affected the results obtained.

In an inquiry such as the present it is important to bear in mind a distinction which previous investigators have in the past too often overlooked. We must carefully discriminate between children who are actually born as twins and those who in any particular investigation may be identified as twin siblings of other children comprised in the group actually investigated. Many of the children who were born as twins may fail to appear in the available records owing to various selection processes—e.g. mortality, separation during the intervening years, or the occasional effects of educational organization.

During the past twelve years, as Chief Examiner or as Selection Officer for the authorities concerned, I have been able to trace more than a thousand twins who have sat in various 11 plus examinations. The main results were given in an earlier paper [6]. Since its publication, however, I have obtained a good deal of supplementary information.

The data analysed here relate to two distinct populations. (1) In one area, with the exception of those officially classified as ' educationally subnormal ', every pupil attending the primary schools was normally expected to take the examination. This will be termed the ' complete age-group '. (2) In the other area only those whose parents wished them to take the examination sat as candidates. This will be referred to as the ' incomplete age-group '. In the second area the final selection was made in two stages, separated by an interval of two months. In both areas, during the years for which full returns

were available, the examination consisted of ' old type ' papers on English and Arithmetic, that is, the questions set and the answers given were longer than those for the ' new type ' examinations now adopted by many authorities in Great Britain.

II. Results

Table I presents the main results. The first line shows the number of (i) twins born in England and Wales in 1955, (ii) those identified as twins in the ' complete age group ', (iii) those identified as twins in the ' incomplete age group ', and (iv) those identified as twins among candidates qualified to pass forward (those termed ' twins ' include siblings who belonged to four sets of triplets). The second line expresses the numbers as proportions of the total populations subdivided into the same four classes. The third line gives the percentages among identified pairs of twins that were of opposite sex. The next three lines show the correlations between the marks of twins according as the pairs consisted (*a*) of boys only, (*b*) of girls only, (*c*) of a boy and a girl. The last four lines show the differences (in standard measure) between the average marks obtained by the twins and those obtained by all the candidates; the figures were computed as follows : from the mean mark of the twins we subtracted the mean mark of all candidates of the same sex, and then divided the difference so obtained by the standard deviation for the marks of candidates of that sex.

Table 1. Proportions, Correlations, and Marks for Twins

	General Population	Complete Age Group	Incomplete Age Group	
			Entrants	Selected Subgroup
Number of Twins	15,414	500	511	110
Proportion per 1000	24·2	16·5	7·9	5·8
Percentage of Twins who are of Opposite Sex	36·3	29·2	29·0	29·1
Correlations for Twins :				
Boys	—	0·85	0·78	0·43
Girls	—	0·86	0·83	0·76
Opposite Sex	—	0·60	0·43	−0·31
Mark Differences (in Standard Measure) :				
Same Sex : Boys	—	−0·57	−0·25	−0·24
Same Sex : Girls	—	−0·47	−0·23	−0·52
Opposite Sex : Boys	—	−0·64	−0·18	−0·20
Opposite Sex : Girls	—	−0·35	−0·13	−0·69

From the size of the samples involved it will be clear that the relevant differences shown in the table are fully significant. The correlations exhibit the usual tendency to become reduced as a consequence of the increasing homogeneity resulting from selection. But, unlike most previous investigators, we have not applied any theoretical correction for this reduction, since it is part of our purpose to draw attention to these effects. Here, too, as in a good many other instances where it has actually been applied, the ordinary formula

would in point of fact be quite invalid. In addition to the selective effects commonly noted, there may often be a further process at work. If, within pairs of twins, there happens to be a decrease in the correlation for survival rates, then, at any specific level of selection and for any individual survival rate, the number of pairs surviving will show a diminution ; and the lower the correlation, the more the mean level of those surviving will tend to fall [6].

The following are some of the more important conclusions to be drawn from an inspection of the figures shown in the table.

1. A comparison of the proportions given for twins in the first two columns (line 2) indicates that the survival rate for twins is less than that for the ordinary population. This difference must almost certainly be due chiefly to differential mortality—a conclusion which agrees with that of earlier investigators.

2. The figures beneath (line 3) suggest that the survival rate for twins of opposite sex is less than that for twins of the same sex. Here, however, part of the difference indicated by the figures available may result, not simply from a differential mortality, but also from an occasional difficulty in identifying twins of opposite sex, especially where they attended schools for one sex only. Certainly the reduction in numbers appears far too large to be attributed to mortality alone.

3. Twins generally speaking show less ability than the other children of the same candidature. This fully bears out the inferences already drawn by Mehrotra and Maxwell [3], who reported somewhat smaller differences, and by Burt and others in a joint study reported in this *Journal* [5].

When we turn to consider the differences in relative ability between *different types* of twin, it seems clear that we can no longer accept the mark differences just as they stand. In the ' complete age group ' the difference between the marks for twins of the same sex and those of opposite sex may be partly due to several somewhat elusive processes of selection. In addition to the differential effects of mortality and of occasional allocation to separate schools, the administrative procedures adopted for identifying the ' educationally subnormal ' may also have influenced the results, since in some cases the effect apparently has been to separate individual twins who are near the borderline. Nevertheless, these influences seem to play only a minor part, and cannot be wholly responsible for the differences shown. As regards twins the chief cause seems clear. As Burt and others have noted, in the human race the very fact of twinning is suggestive of some abnormality in the pre-natal processes, and the conditions of gestation and delivery are more likely to be unfavourable [5, 7]. This would account both for the high mortality observed among twins and for the weaker physique and mental capacities of many of those who survive. Moreover, this interpretation seems confirmed by an observation of Dr. B. H. Kirman, who points out, in a personal communication, that among patients in mental hospitals a feature complementary to our conclusions may be discerned : the proportion of twins found among the mentally deficient inmates is appreciably higher than the proportion of twins found in the general population.

If we now look at the data shown in the last two columns of the table—those for the ' incomplete age group ', giving figures for the two successive stages of the examination—we find evidence suggesting the possibility of yet further processes of selection. To begin with, parents of twins who differ widely in their ability may decide that one only of the pair should enter for the examination. Such a child therefore would not be identified in the official returns as a member of a pair of twins. In consequence, if the correlation between twins of this type is low, the principle stated above will be operative ; and the ultimate effect might be to raise the ascertained marks for twins of this type as compared with those for the children as a whole. But secondly it would appear that the preliminary examination eliminates an appreciably higher percentage of twins than of ordinary children. However, it seems clear that these processes have not operated to any very important degree, since the marks for twins of opposite sex suggest that these do not exhibit so pronounced a defect in mental capacity as do those of the same sex.

4. The correlation between the marks obtained by pairs of twins is higher for those of the same sex than for those of opposite sex. For twins of opposite sex the coefficients obtained appear to be much the same as those commonly reported for brothers and sisters of the same family who are not twins. On theoretical grounds, as Hogben has observed, we should expect that, if all twins were of binovular origin, the number of pairs of the same sex would be equal to the number of pairs of opposite sex ([2], ch. i, ' The Problem of Twin Resemblance ') ; hence the increased frequency of twins of the same sex must be due to the fact that many of them are of uniovular origin, i.e. must be identical twins. According to the figures he has brought together from a variety of sources it would appear that approximately 25 per cent of all human twins are identical. In my previous paper [6], cited above, the results reported are fully consistent with this view. The analysis indicated that, among the twins of the same sex, the numbers of each of the two kinds were approximately equal both at birth and at the age of ten ; and the conclusion finally reached was that the correlation between identical twins was of the order of 0·96. Miss Howard has pointed out that much the same figure is reached by an alternative mode of calculation. Using methods similar to those proposed by Fisher [1], we may estimate that, in the complete group, the correlation for twins of the same sex would be approximately 0·73 (probably rather more in the case of the girls and rather less in the case of the boys) and that the correlation for identical twins would be approximately 0·97. These results are in keeping with the hypothesis that, in the case of non-identical twins of the same sex, part of the increased correlation is due to genetic factors (as with sibs of the same sex), and that part is due to similarity of environment (e.g. children of the same sex may be sent to the same schools, while children of different sex may be sent to different schools). The high correlation for identical twins provides still further evidence for the importance of genetic factors as determinants of ' intelligence '.

I may add that the figures given in the last column of the table—those for the subgroup from which final selection for transference to the grammar school was to be made—are borne out by long personal experience as a grammar school master for nearly forty years. In schools of this type the pupils form a relatively homogeneous population, and it is generally possible to identify, with high probability, those who are identical twins. Such children usually resemble each other in their abilities far more closely than non-identical twins, while twins of opposite sex resemble each other far less than twins of the same sex. Since so many irrelevant factors—notably a wide variety of selective conditions—affect data for twins collected in grammar schools, it is clear that the results obtained from these schools only (as in the case of many earlier inquiries) must be interpreted with far more caution than has hitherto been the case. The data obtained in the present investigation largely circumvents this difficulty.

III. Postscript

After the foregoing paper was set up in type, a good deal of further material became available which seems fully to confirm the conclusions reached above, and—what is more important—strongly suggest that the inferences there drawn are not merely temporary results, due perhaps (as some critics have apparently suggested) to special peculiarities of the examination in its earlier stages, but rather the effects of more or less permanent characteristics of twins as such. I am therefore grateful to the Editors for permitting to append this further note.

The new data relate to the results of the 11 plus examinations for 1958 and 1959, and thus represent the latest figures available. Among just over 184 000 candidates as many as 181 individuals were traced who were survivors of multiple births. This is equivalent to 9·8 per mille, a rather smaller proportion than that cited above in Table I (16·5 per mille). Moreover, the group includes only nine pairs of twins of opposite sex, i.e., 5 per cent of the twin group instead of about 30 per cent, as we should expect. These lower figures are no doubt due to the fact that, during the shorter period of time that has elapsed, it has not been possible to trace alarge number of the twins of opposite sex who would ordinarily go to different schools. Among those who were of the same sex there were 72 twins who were both boys; and for these the correlations between the marks obtained by the two members of the same pair amounted to 0·78 in 1958 and 0·91 in 1959 : the defect in their marks amounted to −0·32. There were 91 who were girls (this includes one case of triplets); and for these the correlations were 0·85 and 0·75 respectively: the defect in their marks amounted to −0·40. Allowing for the size of the sampling errors these figures fully agree with those derived from the earlier data[1].

[1] *Editorial Note.*—Data published by Professor Husén of Stockholm since Dr. Sandon's paper was originally received appear to be in close agreement with the conclusions reported by Dr. Sandon and the other writers whom he quotes. The differences in intelligence between non-identical twins were on an average 1·7 times greater than the differences between identical twins. Moreover, at every age the twins had a lower average I.Q. than the other children: 'Psychological Twin Research', *Stockholm Studies in Educational Psychology.* No. 3. (C.B.)

References

[1] Fisher, R. A. (1925). The resemblance between twins. *Genetics*, I, 569–579.
[2] Hogben, L. (1931). *Genetic Principles in Medicine and Social Science*. London : Williams and Norgate.
[3] Mehrotra, S. N., and Maxwell, J. (1949). The intelligence of twins : a comparative study of eleven-year-old twins. *Population Studies*, III, 295–302.
[4] Burt, C. (1955). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.*, XXV, 158–177.
[5] Burt, C., and Howard, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. Statist. Psychol.*, IX, 95–131.
[6] Sandon, F. (1957). The relative numbers and abilities of some ten-year-old twins. *J. Roy. Statist. Soc.*, CXX, 440–450.
[7] Drillien, C. M. (1958). The survival rate of twins. *Arch. Dis. Child*, XXX, 10.

Vol. XII Part II The British Journal of Statistical Psychology November 1959

STATISTICAL METHODS IN PSYCHICAL RESEARCH

By C. C. L. Gregory
Formerly Director, University of London Observatory

In recent controversies concerning the statistical significance of the results of card-guessing in parapsychology, the possibility of using the guesses themselves to provide a criterion of significance has not been adequately considered. Arguments are put forward in favour of employing for this purpose elementary and obvious methods in preference to the conventional statistical methods which assume a strictly random distribution in the target sequence. A worked example of a simple procedure is described; and a proposal is made to test the so-called P.-K. effect, by means of the 'Galton board' (a pin-table which, in the absence of any such influence, gives a Gaussian distribution).

I. Alternative Statistical Tests

In psychical research so-called statistical methods are supposed to have the special advantage of providing a measure of the probability that the reported occurrences are due to chance alone. If the probability is very small, this would imply either that some form of paranormal causation is at work or else that some events at least are uncaused. Morevoer, it must always be remembered that the statistical probability of any occurrence necessarily depends on a sufficiency of mutually compatible assumptions to determine it mathematically: in this respect the general practice of the statistically oriented psychical research-worker of today is the reverse of that met with in other sciences (with the single exception of what is termed *curve fitting*). Nevertheless it is just this important condition that is so often overlooked. It is perhaps this neglect rather than uncertainty of the validity of probability theory that has given rise to differences of opinion among both professional and lay parapsychologists—as, for example, in the case of Mr. Spencer Brown and Mr. Christopher Scott [1, 2].

Until parapsychologists can discover how to vary the results of their tests by changing the test conditions, their task must remain more like that of an astronomer, who can measure but not experiment, than that of the ordinary psychologist. The practical astronomer is skilful in devising simple statistical techniques, which are often found to have advantages over the more elaborate methods that have tended to become standard in most psychological work. An example is the Dyson method of determining the plate constants of astronomical photographs and their probable errors by minimizing the errors instead of the mean square errors. It is often easier to use the average error, $a = (1/n)\Sigma|x - x_0|$, rather than the standard deviation, σ, where $\sigma^2 = (1/n)\Sigma(x - x_0)^2$.

In curve fitting this has the advantage that large errors do not produce exaggerated distortion. If the distribution is normal, the standard deviation is $\sigma = \sqrt{(\frac{1}{2}\pi)}a$, and the probable error of a single measure is $p = 0{\cdot}8454a$.

When an astronomer determines the distance of a star, in principle he finds the dimensions of a minute ellipse on an ideal photographic plate. This would correspond with a picture of the earth's orbit round the sun as seen from the star, were such a thing possible. If he measures the distance of the same star on another occasion, he will usually obtain a different distance. But this will not cause him to believe that the star's distance has changed appreciably between the two observations; to do so would be to suppose the star had moved relatively to himself with a speed perhaps even greater than the speed of light.

Some stars have had their distances determined by a number of different observers on different occasions with, on the whole, reasonably good agreement. The most probable distance of a star is usually supposed to be given by the mean of all the observations; but if one of these should happen to diverge from the mean by more than about three times the average deviation, it is customary to reject it under the Pierce–Chauvenet rule[1] for rejecting discrepant observations. Likewise, in fitting curves to observations it is customary to choose a curve which minimizes either the mean divergences, or the squares of the divergences from the curve, of the points corresponding to the observations. If a very divergent observation were included, then, when the sum of the squares of the divergences is made a minimum, the inclusion of this single divergent value would be likely to distort the resulting curve unduly; and for this reason it is usual to reject it.

Perhaps the commonest example of a divergence between a probable value calculated by standard theory and one that is a result of measurements is the standard deviation of the mean of two measures. Since $\sigma_{\bar{x}} = (n-1)^{-1/2}\sigma_x$, where n is the number of independent measures, $\sigma_{\bar{x}} = \sigma_x$ when $n = 2$; and consequently in theory there should be no advantage in measuring a photographic plate twice and taking the mean of the measures. Nevertheless that is the standard practice in measuring both star images and spectrum lines for the following reason. In bisecting a short length by eye (for example, in marking the middle of a short strip of paper) it is easy to show by folding that there is a nearly constant bias to one side which is nearly constant for a given width, but which decreases with the width. Consequently a systematic source of error is greatly reduced by using the mean of two measures made in the direct and reversed orientations of the photographic plate respectively.

Statistics are of course used in connection with many biological and psychological experiments to determine how likely it is that a supposed cause is operative; and, if the chance of its being so is as much as 95 per cent, this is usually taken as verification. According to Sir Ronald Fisher [3], "The test of significance only tells him [the experimenter] what to ignore, namely all experiments in

[1][See Editorial Note, p. 147].

which significant results were not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to describe an experiment, so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to produce, are left in suspense pending further investigations ".

What distinguishes parapsychological from other scientific experiments is the lack of an intelligible causal factor. When, for instance, Basil Shackleton, or any other similarly gifted person, has under suitable test conditions correctly guessed six or more cards of a Zener pack out of 25 at least 19 times out of 20, this result should be considered sufficient to establish that some causal factor is operating in addition to chance. If, however, such a causal factor be assumed, and named *psi* or *ESP*, that, without further hypothesis, is commonly treated as the end of the matter. *Psi* or *ESP*, however, are merely names denoting the absence of a hypothesis to explain an effect probably not due to chance alone.

Unfortunately statistical tests of significance have often been used by parapsychologists for quite other purposes than those for which they were originally devised; indeed they have been so used, not only by parapsychologists, but even by some academic psychologists also. Without a hypothesis all that statistics can ever show, in any parapsychological experiment, is a significantly greater degree of similarity between the target events and the guessing events than would be likely to occur to this extent by chance alone. If, however, *psi* or *ESP* is put forward as a causal hypothesis, then some further and different experiments are required to distinguish between this hypothesis and some other hypothesis, as for example, that there is a common causal factor in the order of the target and of the guesses respectively. It is always assumed that the target has been randomized, whatever that may mean, and, in the case of more than chance correspondence, that the guessing is not entirely the effect of chance, but is related in some way to the actual order of the target cards.

In an early paper Soal [4] maintained that Rhine and Pratt's target cards could not form a strictly random selection, because a pack was used which always contained exactly five cards of each suit. Soal himself introduced the practice of using a random selection of cards determined by the order of occurrence of certain numbers chosen according to some pattern from mathematical tables, such as tables of logarithms, where the digits were assumed to be distributed in a random manner. This method was introduced with a view to simplifying the statistical calculations. But I suggest that the only possible advantage to be gained by the use of a so-called random order for the target items is to save the experimenter a little time and trouble in providing a comparable significance test. It has, however, the great disadvantage that it permits the use of statistical devices which are of doubtful validity for determining quantitatively the degree of significance of the coincidences, i.e. a set of coincidences between the target cards and the guesses that agree to an extent greater than chance expectation.

Dr. Soal gave a further reason for using a so-called random target sequence [5]. He suggested that, if the target sequence had a marked pattern or characteristic,

this might be learned by the guesser and used predictively. Clearly, if no serious attempt is made to ' randomize ' the target sequence, it is essential that the guesser should not know by any normal means what was the target order of the guesses immediately preceding.

II. The New Procedure

I shall now describe a method of evaluating the significance of the guesses corresponding to the target Zener cards (or for that matter to any kind of target, e.g. the clock face cards devised by Mr. Fisk) without assuming that the cards which form the target are in a strictly random order. Let us first consider guesses in which Zener cards have been used. If the guess and the target symbol are written in two parallel columns in the usual way, a third column can be added in which a *one* is entered to denote a guess that is correct and a *nought* for an incorrect guess. We may designate the total number of correct guesses $p(0)$. Now let us move each target symbol downwards by a single place, and count up once again the number of the cases in which the target symbol and the symbol guessed happen to coincide. We may call the number of coincidences now obtained (i.e. the total number of correct $+1$ guesses) $p(+1)$. Now suppose the target guesses to have been moved one place backwards: in this case we can call the number of coincidences now obtained '$p(-1)$'. And in this way let us continue the displacements and the calculations until we have obtained all the successive values from say '$p(+13)$' to '$p(-12)$' (cf. table).

In practice these displacements and these calculations can most readily be made if the target sequence is indicated by setting out the 25 target symbols in their actual order at equal intervals round the edge of a circular cardboard disc. The disc must be pivoted at its centre, like a turn-table, to a larger card, which is fixed, and on which signs corresponding to the guesses are similarly arranged in a larger circle. In the initial position of the rotating disc, the target cards will appear opposite the symbols *actually* guessed, i.e. in the undisplaced or zero position. If the number of correct guesses is significantly greater than chance, then the value of $p(0)$ thus calculated should be significantly greater than any other of the other p-values obtained by rotating the card through one place at a time. The artificial values of p thus serve as a basis from which a value for the probability required can be estimated.

The simplest way of performing the calculation is to take the average difference between each p-value and the mean of all the p-values; then, if any p-value is more than three times this measure, it will be deemed discordant according to the Pierce–Chauvenet criterion. So long as a significant discordance with the same $p(x)$ is obtained on about 19 out of 20 occasions, we may reasonably conclude that some factor other than chance has determined the guesses.

Described in this way the operation may appear somewhat elaborate; but in practice it will be found simple, speedy, and reliable. The system is easily modified for use with the Fisk clock cards by using more third column numbers

than 0 and 1. If these additional numbers are used, the result can be shown graphically by means of a histogram; and from this both *displacement* and *significance* can be determined in the usual way.

Instead of the turn-table an alternative device is to continue either the target column or the column of guesses in both directions by repeating series in the same order. If the two columns are on separate pieces of paper, one can be displaced step by step relatively to the other so as to give in all the 25 positions; the number of agreements in each position is then counted as before.

We have taken 25 as the number of guesses per test because that happens to be the number of cards in a Zener pack. It is better, however, to use a rather larger number. Karl Pearson used to give 30 as the smallest sample it was safe to use in similar investigations. As an example we show the application of the method to the first 60 guesses, in two lots of 30, given on p. 142 by Rhine and Pratt in *Parapsychology* (1957, English edition).

CALCULATION OF AVERAGE AND SIGNIFICANT DEVIATIONS FOR 60 GUESSES

From Rhine and Pratt's Record Sheet

G denotes symbol guessed; T denotes symbol on target card

First 30 guesses

G	T	G	T	G	T
∧	∧	+	+	∧	∧
O	L	+	=	+	+
O	L	L	L	=	=
=	=	O	∧	L	O
+	∧	∧	∧	=	+
=	L	L	O	O	L
L	L	O	=	L	∧
	O	=	=	∧	L
	O	O	+	L	O
O	O	+	+	+	∧

Second 30 guesses

G	T	G	T	G	T
∧	=	O	∧	O	=
=	=	=	O	=	O
=	+	∧	O	∧	∧
O	=	∧	∧	+	O
∧	+	+	O	=	+
L	+	L	=	O	=
O	+	+	+	L	L
+	=	=	∧	∧	O
L	O	+	L	L	L
=	L	O	L	L	∧

p (0)	12	+6	p (− 1)	12	+6
p (+ 1)	1	−5	p (− 2)	6	0
p (+ 2)	4	−2	p (− 5)	5	−1
p (+ 3)	4	−2	p (− 4)	5	−1
p (+ 4)	3	−3	p (− 5)	5	−1
p (+ 5)	7	+1	p (− 6)	6	0
p (+ 6)	10	+4	p (− 7)	4	−2
p (+ 7)	6	0	p (− 8)	5	−1
p (+ 8)	6	0	p (− 9)	7	+1
p (+ 9)	5	−1	p (−10)	9	+3
p (+10)	6	0	p (−11)	9	+3
p (+11)	4	−2	p (−12)	2	−4
p (+12)	7	+1	p (−13)	5	−1
p (+13)	7	+1	p (−14)	5	−1
p (+14)	9	+3	p (−15)	4	−2

Average deviation ± 1·9
Significant deviation ± 5·7
p(0) and p(−1) are significant

p (0)	5	−1	p (− 1)	6	0
p (+ 1)	7	+1	p (− 2)	9	+3
p (+ 2)	8	+2	p (− 3)	3	−3
p (+ 3)	3	−3	p (− 4)	4	−3
p (+ 4)	7	+1	p (− 5)	5	−1
p (+ 5)	4	−2	p (− 6)	6	0
p (+ 6)	6	0	p (− 7)	5	−1
p (+ 7)	3	−3	p (− 8)	7	+1
p (+ 8)	5	−1	p (− 9)	8	+2
p (+ 9)	8	+2	p (−10)	9	+3
p (+10)	6	0	p (−11)	6	0
p (+11)	7	+1	p (−12)	9	+3
p (+12)	2	−4	p (−13)	7	+1
p (+13)	12	+6	p (−14)	8	+2
p (+14)	5	−1	p (−15)	1	−5

Average deviation ± 1·87
Significant deviation ± 5·6
p(+13) is significant

III. Psycho-kinetic Effects

In PK tests it is suggested that a different method of investigation might be tried. It now seems fairly clear that no significant long-term overall effects are to be expected, and that the PK effect, if it does exist, is only occasional and sporadic. All PK effects that have so far been claimed as demonstrated by statistical methods have arisen in connection with small falling, bouncing or colliding objects, such as balls or small cubes. In the roulette play at Monte Carlo, a turn-table, having a uniform circular groove divided into equal numbered cells, is spun by hand in one direction, while a small ball is thrown into the groove in the opposing direction. Eventually the ball comes to rest in one of the numbered cells; the turn-table is then brought to rest gently without disturbing the ball, and the number called out. The numbers are subsequently published in the printed lists of all throws, issued by the Casino for the use of their patrons. It seems unlikely that there should be any undetected errors in these printed tables, because many players keep their own lists, and so provide a check.

Seeking a long, naturally occurring sequence of random numbers, Karl Pearson examined the figures of a fortnight's play, and discovered to his surprise, that, had the game been "played continuously until now from remote geological times, it would be unlikely that one fortnight's play should have occurred by chance which diverged so greatly from what was to be expected if the laws of chance were operative". Some years later I heard that two other investigators had found a like result, and had concluded that "roulette, as played at Monte Carlo, is not a game of chance".

At present there seems to be some difference of opinion as to the cause of this apparent departure of roulette play from chance expectation. It has been suggested that perhaps the wheel is not quite perfect, or else that the printed tables are inaccurate. It should be possible to test both these hypotheses; and, if they should be shown to be inadequate, an analysis of the Monte Carlo play might be expected to reveal the detailed workings of the supposed PK effect.

It may be suggested that new experiments should be carried out with a modified Galton board. This is a sort of vertical glass-covered 'pin-table', rigidly supported, and attached to a spirit level of high accuracy [6]. The advantage of using the Galton board would be that the result of any trial, which takes but a few seconds, can be directly compared with the normal frequency curve $y = y_0 \exp[-(x-x_0)^2/2\sigma^2]$, where in the familiar notation $(x-x_0)$ and σ are the actual and standard deviations from the mean, y the frequency, and y_0 a frequency scale constant.

In the physical sciences it has been customary to compare frequency of unobserved measure with that given by this standard frequency curve, sometimes called (by physicists) a *Maxwellian distribution*. When Eddington compared the measured velocities of stars relative to the sun with values so calculated, he found that the stellar motions did not conform with chance expectations; and 'star streaming' was thus discovered by finding what theoretical velocity-distribution best fitted the measures.

In two important but little-known papers [7, 8], published respectively in 1899 and 1902, Karl Pearson applied his mathematical theory of errors to repeated measurements made by a number of co-workers, and discovered that "the errors made by two observers varied sympathetically; . . . in one series Lee and Macdonell showed a high correlation". Thus it might almost be said that Pearson independently discovered *psi* by statistical analysis as a group effect. This group effect was later investigated by Flugel [9] and Philpott [10] and others, but many psychologists have regarded their findings as too unpalatable, or too 'improbable' to be accepted.

A convenient method of testing significance is provided by the well known function chi-squared (χ^2), which Pearson originally devised as a means of assessing 'goodness of fit' in problems of curve-fitting. It may be appropriately used to assess the amount of agreement between observation and hypothesis, whenever the hypothesis is to be tested by counting the number of individual cases, such as instances of agreements between 'guesses' with target symbols. Moreover (what is not so well known) in such investigations as the present χ^2 can be rightly used to test whether or not data can be pooled.

Thus, if a pack of cards is guessed with an average of seven correct when the chance expectation is only five correct, then the longer the good guessing continues, the greater will be the value of χ^2, and of course the more inappropriate it becomes to describe the calls as 'guesses' since such a term implies that chance alone is operative. Viewed in this light, nothing more would be gained by seeking very large values of χ^2, unless, as Spencer Brown has suggested, we are concerned with an immensely large number of possibilities; and that this is certainly not the case in experiments on card-guessing when complete packs of Zener cards are used and the number of possible guesses is limited to five.

IV. Hypotheses and their Difficulties

There is a well-known difficulty which has always appeared insuperable when attempts are made to suggest a possible *psi* process. If telepathy is a process of 'knowing something in some one else's mind', what is the directive? And if clairvoyance is a kind of seeing, how is this directed to the order of cards in an unseen pack? One way of answering these questions would be to suggest that the necessary restriction is provided by instructing the guesser that only one of five specified kinds of guess is to be made on each occasion. Thus, the number of possibilities in the target and for the guesser on each guessing occasion is reduced to five, which makes the chance of successful matching one in five for every guess. If significantly better results are obtained, then this might be due to some new kind of chance, or to some additional piece of information.

We cannot readily suppose it is due to chance alone, for this could only come about if there were a sufficient number and range of chance possibilities for large divergences from a mean value to be possible. In a normal curve of error the frequency of large deviations from the mean for a given standard deviation depends upon the size of the total population. If we suppose the guesser has full

information of the target ordering, but is prevented by fear, or some other cause, from using more than a small and nearly constant fraction of this information, it is difficult to imagine a process of discarding which is so strikingly effective in reducing to chance level and so nearly constant as Soal found it to be in the case of his best subjects.

If we postulate that chance does not enter at all, so long as there is *psi* guessing, then some sort of *filter theory* seems to be required to account for the fact that so many cards are incorrectly named. Most filter theories designed to account for *psi* failure follow the Freudian concept of a censoring process, supposed to account for every *Fehlleistung* [11] (or 'slip', as it is frequently translated).

There would appear to be at least three objections to this theory as anything more than a supplementary hypothesis. First, we have no grounds for presuming that human beings are endowed with anything remotely like omniscience; and clearly a moderate success with *psi*-guessing alone can scarcely be deemed a sufficient ground for so far-reaching an assumption. Secondly, the hypothesis of a 'censor', if adopted, must surely preclude the possibility of making any worthwhile experiments, because these would pre-suppose that the *omniscient censor* could be tricked by the simple devices of the experimenters. If the censor could be removed altogether, we should presumably be left with *faultless guessing* which would be, to say the least, a most unusual occurrence. The third objection to the filter theory is that it really leaves the main effect, i.e. the supposed *psi*-faculty, entirely unexplained. The known gauge of a sieve might explain the *absence* of particles larger or smaller than a certain size; but it obviously could not account for the particles themselves that it permits to pass.

Let us suppose that each guess which is made during a period when *psi* is operative is accompanied by the reception, on the part of the guesser, of just one 'bit' (binary unit) of paranormal information per guess, which takes the form of 'same as' or 'different form', with reference to the similarity between target and the guess. Thus, if the chance guess hazarded by the guesser produces an *image*, conscious or unconscious, which happens to coincide with the target event, then the bit of information will confirm this venture; and the corresponding symbol will be spoken or written by the guesser, but if the information is to the effect that the image and the target card are dissimilar, then another venture might be made. The chance success will now be one in four so that, on the average, five out of 25 first guesses would be expected, plus five second guesses out of the remaining 20. Thus, ten correct guesses would be expected out of 25 unless further '*psi*-scanning' is postulated.

Even on the assumption of one 'bit' of information per guess it is not to be expected that ten correct guesses out of twenty-five will be likely always to occur in practice, because there must be at least one further factor operating which will affect the guessing, namely, a conscious or unconscious attempt to guess, on the average, five of each suit every twenty-five guesses, particularly in a closed pack. This attempt would be expected to clash, at least on occasions, with the bits of paranormally acquired information, and to reduce success from 10 in 25 to a

somewhat lower figure. An examination of the actual sequence of guesses made in such cases often appears to suggest a half-unconscious attempt to make what the guesser might suppose to be a random distribution of guesses.

References

[1] Brown, G. Spencer (1953). *Nature*, CLXXII, No. 4369, p. 154.
[2] Scott, C. (1958). *J. Soc. Psychical Research*, XXXIX, 217–227.
[3] Fisher, R. A. (1929). *Proc. Soc. Psychical Research*, XXXIX, 191–205.
[4] Soal, S. G. (1937). *J. Soc. Psychical Research*, XXX, 55–63.
[5] Soal, S. G. (1937). *J. Soc. Psychical Research*, XXX, 83–91.
[6] Blüh, O., and Elder, J. D. (1955). *Principles and Applications of Psychics*. Edinburgh: Oliver and Boyd.
[7] Pearson, K., Lee, A., and Moor, L. B. (1899). Mathematical contributions to the theory of evolution. *Phil. Trans.*, A, CXCII, 257–273.
[8] Pearson, K. (1902). On the systematic fitting of curves to observations and measurements. *Biometrika*, I, 265–279, II, 1–23.
[9] Flugel, J. C. (1928). Practice, fatigue and oscillation. *Brit. J. Psychol. Mon. Sup.*, IV, xiii.
[10] Philpott, S. J. F., (1932). Fluctuation in human output. *Brit. J. Psychol. Mon. Sup.*, VI, xvii.
[11] Freud, S. (1921). *Vorlesungen zur Einführung in die Psychoanalyse*. 1. Teil, ' Die Fehlleistungen '. (*Gesammelte Werke*, Bd. XI), 7–76. Berlin: Imago Press.

Editorial Note. As many readers may not be familiar with the Pierce–Chauvenet criterion, they will perhaps be glad to have it stated in Pierce's own words: " Discrepant observations should be rejected when (i) the probability of the system of errors obtained by *retaining* them is less than (ii) the probability of the system of errors obtained by *rejecting* them, multiplied by (iii) the probability of making so many (and no more) abnormal observations " (*Gould's Astronomical Journal*, Cambridge, Mass., II, p. 161). A full discussion of the problem, together with Pierce's investigation of it, is given in the Appendix to W. Chauvenet's *Manual of Spherical and Practical Astronomy* (London, Trubner, 1863, II, pp. 558f.). For more recent reviews of the whole subject the reader may consult P. R. Rider, ' Criteria for the Rejection of Observations ' (Washington University Studies, *Science and Technology*, No. viii, 1933) and W. R. Thomson, *Ann. Math. Stats.* VI, 1935, pp. 214–228. It should be added that the original draft of Mr. Gregory's paper was first submitted on 22 January, and the delay in publication is due partly to the strike in the printing trade. Dr. Soal's reply follows on pp. 149–152.

Vol. XII Part II The British Journal of Statistical Psychology November 1959

STATISTICAL METHODS IN PSYCHICAL RESEARCH

Comments on Mr. Gregory's paper

By S. G. Soal
Formerly of Queen Mary College, University of London

In a brief discussion of card-guessing distributions it is argued that the use of the Pierce–Chauvenet rule, as proposed by Mr. Gregory, would not necessarily apply to experiments in which the target series was not even approximately random. Certain objections are also urged against Mr. Gregory's views on the play at Monte Carlo and on the nature of the information received by successful card-guessers whose achievements apparently defy the laws of chance.

Mr. Gregory's paper is of great interest to all engaged in psychical research, particularly in view of his claim that his ' relative method ' for evaluating the results of card-guessing experiments dispenses with any need for the ' target ' series to be a ' random ' series. When a person compiles in his mind a sequence of letters—say a set of 200 chosen from the five letters E, G, L, P, Z (repetition of the same letter being permitted), then it is practically certain that the sequence will not be ' random ', i.e. it will fail to satisfy the usual tests for a Bernoulli distribution with $p = 1/5$. This happens, because, as Borel once said, the human mind with its habits of unconscious preference and its association of ideas, is quite incapable of imitating chance, however hard it may try. And his remark holds good, whether the guesser is attempting to reproduce a series of ' concealed targets ' i.e. an actual but unknown sequence, or whether he is just writing down the letters with no such object in view.

However, in order to apply a test of statistical significance based on the normal probability integral, it is not necessary that the target sequence should itself form a Bernoulli distribution. All that is required is that in n trials the distribution of those trials which yield 0, 1, 2 ..., n correct hits or successes should (provided n is sufficiently large) tend to a normal distribution whenever there is no causal interaction between the targets series and the series of guesses. Nowadays most parapsychologists compile their target lists with the aid of tables of so-called ' random ' numbers, which are assumed to form Bernoulli series. But empirical tests have shown that such tables are not always trustworthy, although the series constructed by R. A. Fisher by taking the last digits from a table of 20 figure logarithms have, over a long period of years, furnished results in excellent agreement with mathematical theory. From the very nature of the concept of ' randomness ', all such tables must, in occasional *isolated sections*, exhibit deviations from mathematical expectation; but such sections are usually marked ' not to be used '.

As is well known, J. B. Rhine employed 5 symbol cards; and his target sequence was so arranged that successive sets of 25 contained exactly 5 cards each bearing one of the symbols. The use of such a non-Bernoulli sequence is perfectly legitimate. The exact distribution of ' hits ' has in fact been worked out by T. N. Greville, who showed that both the mathematical ' expectation ' and the variance were in close agreement with the binomial formulae, except when the number of trials was extremely large. When the number of trials exceeds 200, the distribution becomes approximately normal. In using such a distribution, however, there is always the proviso that the cards in each set of 25 should be thoroughly shuffled. Target sequences can also be made up with widely differing numbers of each symbol; and the expectation and variance of ' hits ' can then be found from the formulae worked out by W. L. Stevens. But here again, since the ' guess ' series is never random, it is essential that the ' target ' cards should be well-mixed.

But provided this precaution is taken, I should not accept Mr. Gregory's contention that the above methods involve the use of statistical devices of doubtful validity (pp. 139–147). Nor can I agree that the ' comparative method ', which he so lucidly describes, is beyond all cavil: for in my view he appears to commit the error of treating the deviations of the individual scores about their observed mean exactly as if they were experimental errors of observation. It is however known from long experience that errors of physical measurement are, to a close approximation, distributed *normally* about their mean; and the use of such a method as the Pierce–Chauvenet rule does in fact implicitly assume the normality of this distribution.

But in a card-guessing experiment we can no longer claim such certainty regarding the distribution of the deviations, particularly when, for example, the list of cards to be guessed is made up in some person's head, and the target distribution is in consequence entirely non-random. Accordingly, since the ' guess ' series is never really ' random ', it would seem that Mr. Gregory is too optimistic in contending that the Pierce–Chauvenet method may be applied, even though the target series may not be random.

On the other hand, in certain cases experience does seem to show that, even when the card-series fails to satisfy some of the requirements of a Bernoulli distribution, nevertheless the expectations and the variance of the direct ' hits ' are (in the absence of ESP or other causal agency) in fair agreement with the binomial formulae. This has been demonstrated by carrying out what is commonly known as a ' cross-check '. Each target column of a successful ESP experiment is checked for ' hits ' against a ' guess ' column chosen from the same experiment for which it was not originally intended. For instance, target-column No. 1 is compared, not with its own ' guess ' column, No. 1, but with ' guess ' column No. 2, and ' target ' column No. 2 with ' guess ' column No. 3, and so on. The total number of hits in this pseudo-experiment is then computed; and the distribution of scores of columns of 25 is fitted to a binomial distribution with $p = 1/5$. From such cross-checks it generally appears that ordinary types of

deviation from randomness in the ' target ' series have little effect either on the expectation or on the variance of the distribution of ' hits '—always assuming of course that the guesser has not been allowed to study the target-lists used in previous experiments. In my personal experience even a considerable deficiency in the number of ' repeats ' (AA, BBB, etc.), or in patterns of the types ABA, has little influence on the expectation and variance of the direct hits, provided there are approximately equal numbers of the different symbols.

If in evaluating a card-guessing experiment we are to ensure that the binomial formulae may be safely employed, then the most important requirements are (*a*) that the numbers of the five ' target ' symbols should be approximately equal, and (*b*) that the cards should first be thoroughly shuffled. If condition (*a*) does not hold, but (*b*) does, then generally speaking, there will be little risk in using the Stevens formulae for the expectation and the variance.

One of Mr. Gregory's dilemmas I should be very loth to accept, viz. *either* (i) that Monte Carlo is a vast emporium for the perennial display of psychokinesis *or* (ii) that in practice chance does not work in accordance with the theories of mathematicians. In spite of the fact that certain players keep private records of the numbers which turn up and compare them with the published lists, probably the majority take no such trouble. It is only too likely, as Whately Carington suggested, that when the Casino clerks become bored, they find it more agreeable to fill in the gaps in their recording from their own imaginations with a glass of beer in front of them. And, to quote Borel once again, no human mind can imitate chance.

But, even if the risks were too great to render such an explanation plausible, there is still the possibility that the roulette wheel and the bouncing ball form altogether too crude a device to generate a sequence of numbers that is strictly random in the mathematical sense. A suitable electronic randomizer would probably be far more successful in generating random series.

Mr. Gregory's discussion in the last part of his paper about the kind of information which enables a psi-subject to beat the laws of chance is to my mind extremely suggestive. I have long suspected that, during a period in which his faculty is working well, a good percipient may, even when he makes a wrong guess, have *some* knowledge relating to the ' target ' card. Supposing that the five possibilities are E, G, L, P, Z, he may for instance know that the card is *not* a P; or he may know that it is neither a P nor an E: he may perhaps run over the five possibilities in his mind one after another, and so get some feeling of certainty that it is not this card or that. Nevertheless, I do not think that the model suggested by Mr. Gregory in his penultimate paragraph is really satisfactory. According to this scheme the psi-subject would, on an average, receive an accurate bit of *positive* information only 5 times out of 25 trials, i.e. on the particular occasions when a chance guess happened to coincide with the card actually envisaged by the agent. When the chance guess happens to be wrong, the percipient receives a piece of confirmatory information to that effect; and this would occur on the average 20 times out of 25 trials. If he then changes

his guess before committing it to paper, his chance of success is now 1 in 4 instead of 1 in 5; and according to this model his average score, even when ESP was active in every trial, would not exceed 10/25 as Mr. Gregory correctly states.

But on certain occasions in the summer of 1956, when the Jones boys, Glyn and Ieuan were at their best, they frequently made in succession three or four scores ranging between 18 and 25[1]. For example, on the morning of 28 August, in the presence of Dr. Donald West, of Mrs. K. M. Goldney and of Mr. Jack Salvin (a sceptical expert on codes and trickery), Glyn, the guesser, sitting behind a large screen with Mr. H. T. Bowden who recorded his guesses, and at a distance of 27 feet from Ieuan the agent, scored 21, 20, 18, 19 correct hits out of 25 trials. And on several other occasions, at distances apart varying from 80 feet to 166 feet, the boys have made similar high scores under conditions that precluded trickery either by themselves or by the regular experimenters.

It is quite obvious that Glyn must have received accurate and exact information about the card many more times than 5 in 25 trials. Yet there is good evidence that, when Glyn was making these high scores, he knew *something* about the card even when his guess was wrong. For instance, in the open-air experiments of August 1956 a 5×5 table shows that, when the target letter was E, he very seldom made the error of guessing P; and, when the card looked at was P, he seldom confused it with E in those cases when he made an incorrect guess. Moreover, these and many similar facts furnish evidence that his phenomena cannot be accounted for by any hypothesis of auditory leakage: for on this hypothesis the similar-sounding letters P, E, G would be likely to cause confusion, and there was not the smallest indication to support this suggestion.

[1] *The Mind Readers* (Faber, 1959). The above considerations would not apply if the requisite information was transmitted by a high frequency whistle concealed in one of the boys' clothes and worked by pressing a rubber bulb. There are, however, numerous reasons for rejecting this explanation. Here it will perhaps be sufficient to cite only one. On 10 August 1956 the boys wore close-fitting bathing costumes and were sitting at a distance of 166 feet from each other. The agent (Ieuan) was carefully searched by myself; and no other member of the family was present. Yet the number of correct guesses was 7, 5, 20, 20, 23, and 20 in 6 runs of 25 each.

NOTES AND CORRESPONDENCE

FIELD THEORIES AND STATISTICAL PSYCHOLOGY

To the Editor, British Journal of Statistical Psychology

SIR,—Towards the close of his article on ' Experiments on Telepathy in Children ' (this *Journal*, XII, p. 73) Professor Burt states somewhat dogmatically that " there is no known hypothesis which will account for such phenomena ", and in a brief footnote apparently sweeps aside the various theories put forward by Thouless, Gardner Murphy, and other investigators, who have sought to explain the facts observed. To fill the gap he puts forward a new suggestion of his own, based apparently on the assumption that individual consciousness is analogous to an electro-magnetic field. This seems to depart as widely from the neurological doctrines of the ordinary textbook as it does from the conclusions of Professor Rhine. Could he therefore explain in slightly fuller detail the empirical evidence on which this conclusion is based ?

E. C. BAKER.

SIR,—In the concluding sections of his review of Dr. Soal's interesting researches Professor Burt appears to formulate two quite incompatible views as to the nature of ' consciousness '—in itself a highly ambiguous term. On pp. 79–80 he tells us that " mind or consciousness " is not to be conceived as a " quasi-physical entity or substance " located in space, but rather as " a unique kind of relation "; on p. 86 he describes consciousness as a " psychic field ", and maintains that " the consciousness of a given individual resides in the environment around him rather than inside his head ", i.e. that it *is* located in space.

If I interpret him aright, the kind of ' field ' he wants us to envisage is not a topological field, like that of Lewin, nor a field of energy like that of Clerk Maxwell, but something resembling the ' psychological field ' described by Whitehead (i.e. the concept which Whitehead himself tells us originally suggested his ' organic conception of reality '). But Whitehead is not the most lucid of writers. Perhaps therefore Professor Burt would explain a little more precisely how he himself is using the term ' field ', and how his conception resembles or differs from those of other writers, and more especially why he supposes it affords an effective explanation of E.S.P. The notion is apparently put forward as a hypothesis to be verified by statistical investigation: what then are the statistical methods we are to apply ? To make my point clearer may I cite a concrete problem ? In playing a rubber of whist yesterday evening I made (to my own surprise) a series of remarkably lucky guesses, not only as to the cards in the other players' hands, but also as to cards they would actually play. How would Professor Burt describe such an experience in terms of his field theory, and what tests of significance would he propose in order to decide whether or not I myself possess clairvoyant or precognitive powers ?

W. R. DAVIDSON.

REPLY

(1) The sentence which Mr. Baker quotes from my article must be interpreted in its context. I gave it as one of the two main reasons adduced by " the majority of psychologists " to explain why in their opinion research on paranormal phenomena appears an " unprofitable pursuit ". In the footnote I added that I myself " fully agreed " with Professor Gardner

Murphy's statement that, " unknown to most psychologists, several coherent theories have been developed ". But the very fact that there are " several " is itself a ground for further research to decide between them; and, so far as I can see, none of them has so far been stated in such a way as to suggest any *experimenta crucis* which would achieve such a decision. The diffident way in which each of the writers puts forward his theory—as a " mere groping gesture to define a process only dimly perceived " (to quote Murphy once again)—implies that they would probably agree with my ultimate conclusion that it would at present be premature to formulate in detail any comprehensive, all-explanatory theory. My argument therefore was that we must proceed piecemeal by testing a succession of more specific hypotheses first of all; and the ' field theory ' was mentioned merely as one of a tentative series.

In comparing consciousness to a ' field ' (in the mathematical rather than the physical sense) I was not putting forward a ' conclusion ' based on ' empirical evidence ' already obtained; still less did I imagine that the comparison was ' new '. I introduced it primarily as a fairly obvious way of meeting the objections raised by Dr. Hansel and others to the assumption of " apparent action at a distance ". At the same time it seemed to offer a possible method of reconciling what I called the ' pluralistic ' hypothesis (which usually treats consciousness or mind as a ' substance ' and relies on causal laws of a quasi-mechanistic type) and the ' monistic ' hypotheses of the semi-mystical writers on paranormal phenomena who seem to ignore instead of facing the many problems raised by the isolation of one individual consciousness from another: in a word, the views of those who treat minds as ultimate and indivisible particles, and those who treat Mind as a kind of all-pervading spiritual medium.

(2) Mr. Davidson's questions call for a rather longer answer. The term ' consciousness ' as Mr. Davidson implies is, like the term ' sensation ', somewhat ambiguous. It may denote a group of *relations* (' subject–object relations ', as they are sometimes called), of which the most typical is the relation of ' awareness ' or direct cognition, as in actual ' sensing '; or it may denote the objective *relata*, the ' contents ' (as I termed them) of which we are directly aware in certain types of cognition and of which the most conspicuous are our so-called *sensa*. On p. 80 I was referring to the ' relation '; and I tried to indicate this by using the phrase ' conscious *of* '; on p. 86 I was referring to what I had called ' contents '. But in discussing a continuous ' field ' the plural would have seemed incongruous. I might have substituted Ward's expressive phrases [1]—' objective *continuum* ' or ' *totum objectivum* ' (which includes ideational and affective contents as well as sensory); this, however, would have been needlessly pedantic; and I assumed that the general context would indicate the meaning I attached to the more familiar term ' consciousness '.

My general hypothesis starts with a broad conception already familiar to most British psychologists and perhaps most clearly stated by Ward. " At any given moment we have a certain whole of presentations, a ' field of consciousness ', psychologically one and continuous; at the next moment we have, not an entirely new field, but a partial change within the old field.[2]" But in accordance with the methodological principles set forth elsewhere (this *Journal*, XI, pp. 31 f.), it seemed to me far easier to test the implications of such a conception if the essential features could be expressed in quantitative terms. My own attempts have therefore been directed towards quantifying both the specific characteristics of the ' field ' at any one moment and the changes in such a ' field ' from one moment to the next.

Mr. Davidson asks how precisely the ' psychic field ' which my hypothesis postulates resembles or differs from the ' fields ' of the physicist. In the physical sciences the term ' field ' is associated primarily with the work of Clerk Maxwell. In his *Treatise on Electricity*

[1] J. Ward, *Psychological Principles* (1918, an expanded version of the well-known article in the 9th edition of the *Encyclopaedia Britannica*), pp. 30 f.

[2] Ward, *op. cit.*, p. 77.

and Magnetism Maxwell wrote: " We may conceive the physical relation between electrified bodies either as the result of the state of an intervening *medium* or as the result of a direct action between bodies *at a distance* "; and his main purpose was to substitute the former interpretation for the latter. The ' medium ' (or at least that part of it in the neighbourhood of the two bodies) he conceived to be in a ' state of stress '; and, availing himself of Lagrange's device of generalized coordinates, he found it possible to describe the state of the medium and its changes by certain partial differential equations. The medium itself was identified with the aether. However, with gradual acceptance of the arguments advanced by Einstein in his celebrated paper (' On the Electrodynamics of Moving Bodies ', 1905) the medium as such has faded away; and the expression ' field theory ' has come to designate a highly abstract mathematical theory describing physical effects transmitted continuously in space (or space-time) in accordance with certain so-called field-equations.

In psychology ' field theories ' of course are by no means new[1]. But (except where the reference is to physical ' fields ' associated with the brain) the ' field ' that is postulated still seems to be conceived as a permanent and quasi-substantial ' medium ', a kind of psychic aether, existing in ordinary three-dimensional space, with properties described in purely qualitative or verbal terms. As will be obvious from my paper and the references there quoted, I myself used the term in the stricter and more abstract sense, analogous to that of Einstein or Whitehead rather than to that of Maxwell. In this narrower sense the concept no longer needs to postulate (though it does not exclude) any concrete hypothetical medium. The model envisaged is a *purely mathematical model*, not a quasi-physical or quasi-substantial model. It is in short the grin without the cat.

The definition I have usually adopted is the following: " the term ' field ' signifies a region within which every point can be associated with a definite value of a certain quantity which in general varies continuously from place to place[2] ". The ' points '

[1] As descriptive terms of psychology such phrases as ' field of vision ', ' field of attention ' (*Sehfeld, Blickfeld,* etc.) were first given currency by Wundt; but the underlying notion is at least as old as Herbart and is hinted at by Leibniz. The concept of a *metric* field (analogous to the electrical and magnetic fields of the physicists) ' interacting ' and ' isomorphic ' with other fields was introduced over forty years ago by the Gestalt psychologists—notably Wertheimer, Köhler, and Koffka. A fairly systematic account of these earlier uses will be found in B. Petermann, *The Gestalt Theory and the Problem of Configuration* (1932; see index s.v. ' field ', ' isomorphism ', etc.). As he points out, " the great idea of Köhler's was to assure the position of the new category of Gestalt by resorting to physics, which has a theoretical apparatus so much better established ". Accordingly, in order to describe the " dynamical distribution systems "—exhibited, for instance, in the phenomena of visual perception—Köhler proposed that they should be " represented in the nomenclature of theoretical physics—by ' fields ', ' lines of force ', etc." Petermann himself is highly critical of these suggestions, at any rate in their original form; and with many of his objections I should agree. Largely as a result of these criticisms, the term ' field ' has come to be used by psychologists in a much broader and more general sense, viz., simply to emphasize the fact that " complex organized wholes cannot be fully understood in terms of ingredient parts " (Gardner Murphy, *J. Amer. Soc. Psych. Res.*, XXXIX, 1945, pp. 181 f.). The word has thus become little more than a convenient monosyllabic synonym for the phrase " unitary structural whole " (*loc. cit.*, p. 196).

[2] A. Haas, *Introduction to Theoretical Physics* (1924) I, p. 33. Lindsay words it slightly different: " it is convenient to treat the region surrounding a particular body (a magnet, a charged particle, or a body with a certain mass) as a kind of continuum in the sense that another body placed at any point therein will be influenced by the first to a degree which in general varies from point to to point: such a region is termed a *field* " (*Concepts and Methods of Theoretical Physics*, 1951, p. 280). Regarded as a vector field, the field of gravity surrounding a given body is characterized by a certain vector defined for every point of the field at every instant of time, viz., the force on a unit mass; regarded as a scalar field it is characterized by the values at those points of the corresponding potential (potential energy per unit mass). Thus distribution of density, temperature, and the like are commonly treated as scalar fields; the flow of fluids, of heat, etc., require treatment by vectors.

(or ' places '), of course, need not be ' points ' in ordinary three-dimensional space. The term ' space ' is merely a convenient way of describing any kind of manifold. Often, however, in psychology it does seem to be closely related to ordinary three-dimensional space (e.g. in the case of visual fields).

In the simplest type of field, the number representing the quantity will be a scalar; it may, however, be a vector (e.g. the gradient of a scalar field at the point in question), or a matrix, or a tensor, or even some still more complicated mathematical expression[1]. In physics the structure of any given field is (in theory) determined by placing a ' test body ' at every point (or at a sufficient number of points)—e.g. with a magnetic field a small compass needle to determine a direction and unit north pole to determine strength.

In what I called ' psychodynamics ' many of the basic problems arise out of the *changes* in the field; and the relevant changes can most simply be expressed in terms of the differences between such scalars. The determination of observable values for the scalars forms the subject of a branch of quantitative psychology which may conveniently be called ' mental campimetry '. In dealing with a cognitive field the crucial variable is obviously the degree of ' cognitive clarity ', i.e., what Wundt termed *der Klarheitsgrad*. This he defined as " the sharpness of the apprehension of conscious contents " (die Schärfe der Auffassung psychischer Inhalte "). Clarity as thus defined is an essential attribute of every cognitive content; and, as he points out, forms the objective aspect of what on the subjective side we usually refer to as the strength or degree of attention[2]. In principle it should be quite possible—though in practice it appears by no means easy—to measure the degree of ' clarity ' for different points (say) of the visual field of perception either by adapting the psychophysical methods used for measuring the intensity of a sensation or (more roughly) by employing a rating scale. Thus, a bright point of light appearing suddenly in the field of vision will in general attract attention, and this will ordinarily be indicated by a movement of the eye. The strength of the attraction can be roughly assessed by the speed of the reaction, and its direction by measurements on a graduated perimeter. (The experiment is cited as an illustration only, since as ordinarily conducted it is as much a study of the field of vision as of the field of attention.) Similarly, it is possible to determine the clarity of different items in the field of visual imagery by the methods used by Galton for assessing ' vividness '[3].

When we attempt to carry out such measurements, we find that the scale of clarity is itself a continuous scale. At any one moment the maximum in the actual range of clarity is located at a single definite point; but there is no very definite minimum. The focus and fringe of ' clear ' attention shade off gradually into a region of inattention; and this in turn into what may be called the marginal area of ' subconscious ' presentations—in one of the many senses of that ambiguous term, namely, in the sense in which the prefix ' sub- ' has the sense ' -ish ' (as in *subruber*—' reddish '); and these dimly conscious presentations themselves seem to fade obscurely away into ' potential presentations ' that are

[1] Cf. Haas, *op. cit.*, chap. iv.

[2] W. Wundt, *Grundzüge der Physiologischen Psychologie* (1908) I, p. 541. I hope it is not too fanciful to point out that the metaphor underlying the traditional term ' attention ' (literally, ' stretching towards ') is that of stretching a bow or the elastic cords of a catapult, when aiming at a mark. In mathematical physics the stress which produces such a stretching is expressed by a ' tensor ' (i.e. a ' stretcher ')—a special type of matrix. To avoid the notion of action at a distance Newton's contemporaries imagined that the ' gravitational field ' (as we should call it today) contained invisible stretched springs, pulling like equally stretched pieces of elastic attached to the central body, and radiating in all direction: in virtue of Hooke's law this model very neatly yields the Newtonian law of inverse squares. A similar model suggested to Faraday his notion of ' lines of force ' making up an electric or magnetic ' field '.

[3] Early attempts in this direction are described by Titchener in his *Experimental Psychology* cf. esp. vol. I, pt. ii, pp. 186–211, and the numerous references. He himself tentatively formulates seven ' laws of attention ' which he claims can be experimentally verified.

' subconscious ' in the sense of being definitely *below* the threshold of consciousness (here the prefix has the meaning of ' beneath ', as in ' submarine '). These ' presentations ' (if with Ward we may still use the term) are presumably not *actually* present at the moment in question, but can be summoned at will into the higher levels of clear consciousness. They thus correspond to the " ideas existing in the *foreconscious* " in the language adopted by certain psychoanalytic writers. Finally, we may, for convenience, assign negative values to those potential presentations which *resist* the individual's efforts to call them up into the higher levels of consciousness—ideas which are in Freudian terminology, repressed into the " unconscious "[1].

The introspective methods I have indicated for describing the cognitive field at any one moment can be prolonged or repeated so as to describe a series of such fields at successive moments. Apart from certain important exceptions the changes will in general be continuous. Our ultimate aim is to determine the type of transformation that is most appropriate for describing such changes. If the course of the changes obeys certain definable laws, the laws will presumably state probable rather than determinate tendencies. They must therefore be expressed as ' transitional probabilities '. And the type of transformation required to formulate the transitions will determine the particular type of field-model we adopt[2].

Although the fields that are familiar to the majority of psychologists are for the most part fields in some kind of *visible* space, it is important to bear in mind that the word ' space ', as used in the present context, does not mean a vacant locus or container in which objects appear to move, but a collection or set of objects (usually called ' points ') which exhibits a definite relational ' structure '. When it is possible to define for any and every pair of such points a function satisfying the conditions for what is called ' distance ', then the space may be called a metric space. For mental qualities ' distance ' commonly expresses the degree of dissimilarity between two things.

For two fields to interact a necessary condition will be that both shall form part of a larger field. Thus my tactile and kinaesthetic fields can interact with my visual field because they are all part of my total field of consciousness. A stranger may sit down with me at dinner; and, if his ideational field largely overlaps with mine because we are both interested in the same subject, then his field may interact with mine. Whether they do in fact interact will depend on the transitional probabilities. If we exchange neither words nor glances, they will not interact—at least under normal circumstances. If they interact in virtue of some ' paranormal ' telepathic rapport that (so I should maintain) is because

[1] Ward explicitly adopts the view that the changes in degree of clarity are continuous, and contends that " the hypothesis of a subconsciousness is nothing more than a further extension of this law of continuity " (*op. cit. sup.*, pp. 90 f.). In experimental work the paradoxical assumption of potential presentations existing below the threshold of consciousness often proves a convenient device. Readers who dislike it can place the absolute zero of their scale at the threshold itself, and reject all negative values. It is curious to note that most early writers speak as if clarity varied only by two or more fairly sharp steps. Baldwin's diagram (*Handbook of Psychology*, 1890, pp. 68 f.), which at one time was so frequently reproduced, distinguished five definite levels, marked off by as many concentric circles. Titchener (*The Psychology of Feeling and Attention* 1908, pp. 220 f.) discusses at some length the question " how many degrees or levels of clearness may co-exist in the same consciousness ? " and concludes by laying down a ' law ' that " there are never more than two degrees ".

[2] The first step will be to define the kind of ' space ' we propose to use to describe the ' manifold ' we wish to investigate. To do this we shall proceed by formulating a series of axioms. The initial axioms will be stated in terms of a class of coordinate systems. We can then deduce the most specific properties of the space from the nature of the transformations of the coordinates permitted by the axioms selected. According to the nature of the data available, we shall add more axioms until we obtain a mathematical model sufficiently isomorphic with the data observed to enable us to answer the questions proposed.

the two fields are, as before, part of a wider field: what exactly this latter field may be, and what conditions are sufficient to provoke such interaction, are problems for further ' psychical research '[1].

My reasons for supposing that a hypothesis such as I have outlined here could claim some measure of *a priori* probability as an explanation of paranormal phenomena were as follows.

(i) The item which is the subject of the paranormal experience (e.g. the object, person, situation, or feeling which is apprehended apparently by clairvoyance, telepathy, precognition, or the like) commonly presents itself to the ' percipient ' as a thought occurring with a relatively *high degree of clarity*. He says: " The idea suddenly struck me "; " I had an unusually vivid feeling "; or (if the thought has concrete form and almost hallucinatory clarity) " I saw him [or it] quite plainly ". When he compares such experiences, it is on the basis of their varying clearness rather than their varying certainty[2].

(ii) The item usually appears in the percipient's *ordinary cognitive field*—his field of vision, his field of imagery, his field of thought. He says: " I saw him in this room ", not " I saw him on his deathbed in New York "; or " I suddenly thought of the three of diamonds ", not " I saw the three of diamonds on the table in front of Harry Price " [the transmitter]; or " I imagined him lying in a pool of blood, and couldn't get the picture out of my mind ". As Gilbert Murray says of his own experiences, " It is more like sym-pathy or co-sensitivity than tele-pathy ". For this reason telepathy (feeling

[1] The general conditions under which a group of individuals tend to form " an organized system of mental forces ", i.e. what is here called a ' field ', has been fully discussed by McDougall in his *Group Mind* (1920, esp. pp. 23 f.; cf. also *id., Psychology: the Study of Behaviour*, 1914, pp. 169 f.). To explain the facts observed several writers have suggested the concept of a " collective mind ". In my view such a phrase may be misleading: a group of trees may form a single wood, but we should not term such a group a " collective tree ". However, as I have elsewhere pointed out (*Scientific Thought in the Twentieth Century*, ed. A. E. Heath, 1951, p. 312) F. Y. Edgeworth long ago " introduced the valuable conception of a psychic ' field ' of force, in which the individual or individuals are set " (see, for instance, *Mathematical Psychics*, 1881, pp. 13, 18, 90); and derived provisional equations to describe such a field by following the methods adopted in Clerk Maxwell's *Treatise*. The notion of a collective psychic field (in Edgeworth's sense) therefore seems to me a far more attractive concept than that of a collective ' mind '.

A somewhat similar notion has been put forward by several recent writers. In a highly interesting paper (' Field Theory and Survival ', *J. Amer. Psych. Soc.*, XXXIX, 1945, pp. 181 f.) Gardner Murphy quotes Lewin's *Dynamic Theory of Personality* as showing that " field theory has much to offer in psychology ", and then goes on to lament that most psychologists, " being ardent individualists, have proved content with the application of field principles to the *individual;* they have not yet learnt . . . that there are instances in which two (or more) individuals may constitute aspects of *one wider indivisible field* ". And in the same journal (XXXVIII, pp. 72–82, ' A Topological Approach to Parapsychology ') Mrs. M. P. Reeves, after showing how " Dr. Murphy has laid the foundation for this (the topological) approach by the use of terms and concepts common to Lewin's topology ", proceeds to illustrate such extensions of the theory by the aid of new Lewinian diagrams. If I understand them rightly their conception (which I should not wish in any way to repudiate) differ from the one I myself have here outlined in several ways: first, it seems to be an environmental rather than introspectable field; secondly, Dr. Murphy and Mrs. Reeves are concerned solely with " non-quantitative relations " (*loc. cit.*, p. 72); thirdly, it is not quite clear what corollaries follow from the hypothesis which would allow it to be empirically tested, and so verified, falsified, or rectified. However, Dr. Murphy's purpose was rather different from my own, namely, to consider how far a ' field theory ' would compel us to modify the hypothesis of *personal* survival: (cf., however, his recent book, *Human Potentialities* 1958, pp. 306 f., on ' The Lewinian Life-space ').

[2] The phrases quoted are from introspective comments of my own subjects or acquaintances: they do not necessarily relate to experiences that were actually verified.

something which is distant), clairvoyance (seeing something in some other place), thought-reading (reading the thoughts in someone else's mind)—all these terms, as commonly understood, seem to be misleading.

(iii) The occurrence of such presentations depends more immediately on the conditions in the *percipient's* own mind and only indirectly on conditions in the transmitter's mind. Hence it is primarily the percipient's field of cognition at the moment of the experience (and at the moments just preceding) that primarily calls for study.

(iv) A relatively *low degree of clarity* as regards other thoughts, especially irrelevant or distracting thoughts, appears usually to form an important pre-condition in the minds of both percipient and transmitter. Greatest success seems usually to be obtained if the transmitter concentrates his attention on the item and so gives it maximum clarity, and if the percipient ' makes his mind a blank '. These conditions, as several investigators have noted, are most readily secured if both are in a state of mild hypnosis.

(v) The thought that presents itself to the percipient seems to exhibit both a certain *strength* (or 'clarity') and a certain *directional relation*. It is related asymmetrically to the thought in the transmitter's mind, and this may be most conveniently treated as a ' direction '. It is natural to express such a directional relation in ordinary spatial and even causal terms. But this is, I believe, to be going too far. When (referring to a normal experience) we ask a man " Where did you get your information from ? " he will reply (if in fact he remembers) " From such a book or from such an event I saw last week ". Here the origin or ' direction ' is not to be interpreted as causal or spatial. So too with interpersonal experiences: the directional relation is not causal or spatial, but epistemological or psychological, though it need not be conscious.

Mr. Davidson asks me to explain how my procedure would be applied to determine whether in the case of (say) a game of whist the accuracy of his guesses were suggestive of " some kind of telepathic rapport with his partner ". In practice this particular problem could be attacked far more easily by means of the usual statistical procedures. However, because of its very simplicity it provides an excellent example for illustrating the hypothesis I have outlined.

Consider then any simple card-game played with ordinary cards. The relevant items of information can be represented by points in a two-dimensional space: we can plot the *values* (ace, 2, . . . king) along one axis of coordinates, and take the other axis to represent the *suits*. At any moment during the play, e.g. when I am about to play the first card, my attention will be focused primarily on one particular card which will have maximum clarity. Probably the other cards in my hand which belong to the same suit will have a fairly high degree of clarity; and the remainder rather less; those in the hands of the rest of the players, so far as I think of them at all, will as a rule have less still. Such a ' cognitive field ' could be represented by assigning numbers representing the degree of clarity to the various points in the two-dimensional space.

A moment later, when (say) the next card is being played, my field of attention will be changed. And throughout the game there will be a succession of changes in the fields of all the players until the game terminates. In the meantime, while the game is still in progress, the field of my partner, as he watches my play, will become more and more isomorphic with my own: if he possessed perfect telepathic rapport with me, his field could be made perfectly isomorphic with mine. Under normal circumstances, however, a good player would be able to predict with increasing accuracy, from his own cards and those already played, how my field of attention is likely to be constructed and to develop as time goes on. The likelihood of such changes he could express by a set of ' transitional probabilities '. An ideal spectator who knew the distribution of all the cards in all the different hands would be able at the outset to work out the successive strategic steps required to maximize my chance of winning. If my actual play came very near to this ideal sequence, he would suspect that I had some exceptional means of securing information

about the distribution of the cards, either by cheating or by some form of clairvoyance or telepathy. Whether the observed results were sufficiently remarkable for him to be able to exclude sheer coincidence as a possible explanation would be determined by the ordinary tests of statistical significance. In short the general plan of the statistical analysis would be much the same as that outlined in some of my earlier investigations of the learning process[1].

Mr. Davidson goes on to ask how ' Professor Burt's theory ' differs from the field theories proposed by other psychologists. It was not my intention to put forward a different ' theory ' of my own as superior to those my correspondents mentioned, but merely to indicate, in a way which might arouse discussion or actual trial, a type of hypothesis which I thought could be formulated in such a form as might render it amenable to statistical verification. In the past, as we have seen, several types of model, based on field theory, have been proposed by early psychologists. Of these the most familiar at the present time appears to be the *topological field* proposed by Lewin[2]. " Topology ", so Lewin reminds us, is " the non-metricized science of relationships " : it investigates those properties of spatial configurations which remain unchanged under transformation, namely, position and connectivity. " In spite of the work of Fechner, Binet, and their followers on what is called psychometrics ", we are told, " the measurement attempts of modern psychology have proved sterile " ; nevertheless, there are many " non-metricized dynamical concepts which are of value in psychological investigations ". D'Arcy Thompson was the first to suggest the use of topology for biological problems—particularly those of morphogenesis; and, since psychology is itself a branch of biology, it was natural to extend this suggestion to the problems of mental life. Except for Lewin's own pupils comparatively little use has been made in actual research of the various methods he himself suggested. Twenty-five years ago topology was not sufficiently developed to furnish an effective instrument for psychological research. Whether metricized or non-metricized methods are to be used depends on the kind of question the investigator is asking. For preliminary tests quantitative methods are to my mind likely to be far more conclusive[3].

The remaining types of field theory proposed by psychologists appear to assume a metricized field of the kind familiar in modern physics. As a result of the progressive unification of electromagnetism, optics, and mechanics carried out by various investigators from Faraday and Clerk Maxwell onwards, almost the whole of modern physics has become a kind of metricized field physics. It was therefore tempting for psychologists to try and

[1] See this *Journal*, IX, 1956, pp. 49 f. Here the state of the individual's ' psychophysical disposition ' at each successive trial can be regarded as the specification of a ' field ': the method of estimating the ' transitional probabilities ' for the cumulative transformations of the ' field ' would be that indicated on pp. 56 f.

[2] K. Lewin, *A Dynamic Theory of Personality* (1935) pp. 117 f., and *Principles of Topological Psychology* (1936). A clearer and more systematic account of the Lewin proposals will be found in J. F. Brown's *Psychology and the Social Order* (1936); see especially the appendices on the methodological background of field theory and on topological and non-metricized variants. Gardner Murphy has applied Lewin's concepts to the problems of paranormal psychology. On the other hand Allport has criticized the whole theory rather severely on the ground that it is concerned " only with the person as he is behaving in some momentary situation " and that it is in consequence not really a " dynamic theory " (*Personality*, 1938, pp. 364 f.). It was my hope that by determining the values for a succession of fields and the transitional probabilities derived from them this objection might largely be overcome (see this *Journal*, IX, pp. 49–50).

[3] A more efficient treatment is to be found in contributions published under the name of Nickolas Bourbaki (*Théorie des Ensembles*, and *Topologie générale*, 1951, A.S.E.I. nos. 1141 and 1142). The basic tool is in my view the general theory of sets, and the choice of the metricized or non-metricized branches for any particular problem must be decided by the type of answer required and type of information available. In practical work a qualitative answer may be all that is needed; if so, a vast amount of labour may be saved if qualitative rather than quantitative arguments are employed. In other cases non-parametric methods may prove most appropriate.

adopt the same concept in dealing with their own particular problems. Edgeworth, as we have seen, applied it as early as 1881 to problems of group psychology, notably those of economics. Köhler applied it to describe the cognitive function of the cortex; Koffka to describe the ' behavioural environment '. And about the same time all references to cognition and to conscious contents went out of fashion. More recently, however, " the term ' cognitive process ', after virtually disappearing from the vocabulary of psychology, has been reappearing " [1], and there has been an increasing attempt to describe or formulate what are designated ' phenomenological fields '. Since the ' field of consciousness ' is itself a ' phenomenological ' field, the field theory I had in mind really belongs to this newer group. It differs chiefly in relying more upon evidence obtainable through introspection[2].

Postscript. Dr. G. D. Wassermann has recently drawn my attention to his paper (read at the CIBA Symposium in 1955) giving ' An Outline of a Field Theory of Organismic Form and Behaviour ' [3]. In his view the hypothetical fields must first be quantized; hence, as he points out, the original feature in his theory arises from the fact that it is based on the ' *quantum* theory of fields '. I should readily agree that both statistical and physiological psychologists may have many things to learn from some of the techniques and conclusions of quantum mechanics; but I doubt whether these include the highly elaborate mathematical procedures used in quantum field theory. So far as I can see, the four special ' properties ' mentioned by Dr. Wassermann could be derived from the more general principles of quantum mechanics without translating those principles into terms of quantized fields[4].

[1] R. Leeper, ' *Cognitive Processes* ', ap. S. S. Stevens, ed., *Handbook of Experimental Psychology*, 1931, p. 736; see the references cited there and on the following pages.

[2] As originally propounded, my own application of ' field theory ' had a further object in view. It was part of my intention to substitute, in current descriptions of mental processes, the concept of psychic fields for the concept of psychophysical dispositions adopted by Stout, McDougall, and others. For McDougall these ' psychophysical dispositions ' were primarily physical, or rather physiological. He pictured them as consisting in systematic patterns of neural channels, formed by groups of neurones linked by ' open synapses ' : J. Clerk Maxwell Garnett termed them ' neurograms ' (a word coined by Morton Prince, *The Unconscious*, 1913, p. 131) and believed them to be amenable to mathematical investigation by factor analysis. Both writers assumed that the neural pattern would be almost completely isomorphic with the psychic pattern exhibited by the individual's conscious experience. The evidence and arguments of Köhler and other Gestaltists showed this neurological theory to be untenable, at any rate in the simple form originally proposed. In any case, we can only identify this or that hypothetical ' neurogram ' by reference to its supposed psychical correlations. For Stout, indeed, the psychophysical dispositions are primarily psychical. In that case the reader is tempted to inquire what kind of ' psyche ' has dispositions that are ' psychical '. Accordingly, to avoid postulating a kind of psychical ' medium ' or ' substance ' with a particular structure, changing from time to time, my suggestion was that we might regard the dispositions as dispositions of a field which could directly be specified in psychic terms—' phenomenal ' terms in Köhler's language—i.e. in terms of ' mental concepts '. However, in this form my hypothesis would require modification or at least re-statement; but this would hardly be relevant here.

[3] *CIBA Foundation Symposium on Extrasensory Perception*, pp. 53–72. I am much indebted to Dr. Wassermann for clearing up several points in correspondence.

[4] Cf. this *Journal*, XI, pp. 75–93. I should agree that to obtain a satisfactory deduction of the " facts and laws of *physics* ", particularly those referred to by Mr. Hansel, Dr. Taylor, and others, something like the quantum theory of fields is desirable (as indeed was implied in my paper); but even this version of quantum mechanics is still in a state of transition. The main effect of the quantization (it is claimed) is to endow the relativistic field with properties that can be expressed by whole numbers, and thus to allow the reappearance of particles. Before it is quantized, such a field is not associated with the concept of particles (see A. March, *Quantum Mechanics of Particles and Wave Fields*, 1951, esp. pp. 179 f.).

But perhaps it is desirable first of all to consider what the transition from the so-called classical field theory to the quantum field theory actually entails—a point which Dr. Wassermann does not, I think, make clear to the ordinary psychological reader. In a paper published some twenty years ago[1], I drew attention to the various ways in which the so called ' new quantum theory ' appeared to be relevant to the problems of the psychologist. In particular I noted that the problems of what in quantum mechanics was commonly called ' spectral analysis ' seemed closely analogous to the problems of ' factor analysis '. For example, the particular factorial procedure I myself considered most effective (based on a technique originally developed by Edgeworth and Karl Pearson) turns essentially, like the procedure commonly used in quantum theory, on the determination of *Eigenvektoren* (latent vectors) and a diagonal or ' canonical ' matrix of *Eigenwerten* (latent roots). Now in the ' quantized ' type of field theory, originally developed by Heisenberg and Pauli (*Zeitschr. f. Physik*, 1929, LVI, pp. 1–61, LIX, pp. 168–190), the canonical variables of the ordinary field equations are replaced by Hermitian operators, p_j, q_j, obeying the commutation rules:

$$[p_j, q_{j'}] = [q_j, q_{j'}] = 0; \quad [p_j, q_{j'}] = (h/2\pi_j)\delta_{jj'},$$

where h = Planck's constant, $\delta_{jj'} = 1$ if $j = j'$ and 0 if $j \neq j'$, and $i = \sqrt{-1}$. This in turn entails the appearance of an *infinite* set of operators which commute with each other. From a mathematical standpoint it is comparatively simple to deal with equations involving the simultaneous spectral representation of a *finite* number of commuting operators; but the introduction of an infinite set leads to highly intricate mathematical problems[2].

Now whereas the matrices (or ' operators ') used in quantum mechanics are Hermitian, those used by the psychologist in factor analysis contain real numbers only. No doubt, as I observed in my paper, " the real axisymmetric matrices—formed, for example, by square correlation or covariance tables—could be regarded as special cases of Hermitian matrices ". Nevertheless, the use of complex numbers in psychological analysis is extremely rare: as a rule it seems to point to some inadequacy either in the data or the methods (cf. this *Journal*, IV, pp. 17–20, VII, p. 26). And I confess that I find it difficult to see how a field theory involving complex numbers could have any intelligible interpretation in psychology. In any case, the mathematical procedure would be extremely laborious.

And generally, Dr. Wassermann's approach to psychological problems diverges widely from my own. His standpoint, as he insists, is " thoroughly behaviouristic ". Accordingly his aim is to " construct a *purely physical theory*, in which mental experiences occur only as epiphenomena " (my italics). And he believes that the " new theory " which he himself has developed will " show that the empirical generalization of embryology, psychology, and parapsychology can *all* be deduced from the *same* initial field theoretical class relations and rules of quantization as the empirical laws of physics " (*loc. cit.*, p. 55, my italics). Now it is an ' empirical generalization of psychology ' (and one with a very high degree of probability) that " all normal human beings feel pain, as I feel pain, when their skin is burnt ". I myself, however, would contend that it is manifestly impossible to " deduce from the field relations and rules of quantization " any proposition which includes the concept of pain or any other content of consciousness, and generally I hold (as will be obvious from the discussions in my paper) that mental experiences are by no means ' epiphenomena ', and that they therefore necessitate a description and an analysis in terms of a ' psychic theory ' rather than a ' purely physical theory ' (cf. this *Journal*, XII, p. 77).

Furthermore, the hypothesis I very tentatively proposed postulated only *one* new type of field, namely a ' psychic field '; Dr. Wassermann's hypothesis, on the other hand,

[1] ' The Unit Hierarchy and its Properties ', *Psychometrika*, III, 1938, pp. 151–168.
[2] Cf. K. O. Friedrichs, *Quantum Theory of Fields*, 1952, and refs.

postulates a large number of different ' quantum theoretical fields ' (in addition to those recognized by writers on quantum mechanics), e.g. various " Lorentz-invariant morphogenetic fields " which he calls M-fields, various types of " behaviour fields " termed B-, B_m-, B_e-fields, certain " specific energy fields " called P-fields, and finally " ψ-fields " which have " narrowly spaced energy levels and occupy wide regions of space ". All of these, we are told, " belong to the same class as the already familiar energy-fields of physics ".

Dr. Wassermann's interpretation of such terms as ' isomorphism ', ' interaction ', ' transitional probabilities ', and the like (as indicated in his preliminary statement of basic assumptions, *loc. cit.*, pp. 54–5) seem to diverge a good deal from the meanings commonly given to them in psychological writings. My own interpretations were designed partly to fit in with the requirements of ' communication theory ' (cf. this *Journal*, *loc. cit.*, p. 89 and refs.). The concepts I had in mind were much the same as those defined in Ross Ashby's work on cybernetics, except that his definitions would have to be generalized so as to apply to a continuum and not merely to discrete elements[1]. It was my view that in many problems of psychology, and certainly in those of telepathy, precognition, and clairvoyance, it might be more helpful to think in terms of the transmission of ' information ' rather than of ' energy '. This incidentally would seem to be the most effective way of dealing with what Jung has called ' synchronicity ', which he describes as an ' acausal principle ' based on connections due to meaning and not to energy or force[2]. In contrast to all this Dr. Wassermann's account of ' interaction ' and ' isomorphism ' is essentially causal. Telepathy, for example, he expressly attributes to the radiation of energy over long distances in space (p. 66). My interpretation, on the other hand, (in terms of a ' super-field ') was put forward as an *alternative* to any form of ' radiation hypothesis ' (p. 75): it would entail (*a*) no time lag and (*b*) no attenuation; and these two corollaries would seem to lend themselves to an empirical test for deciding between the two main alternatives. I may perhaps sum up the difference between us by saying that the purpose of my field theory is to reinstate the study of consciousness, and of Wassermann's to exclude it.

However, it is still conceivable that Dr. Wassermann's ' new theory ' may prove of interest to the physiologist. The fact that a mathematical model involving complex numbers has no apparent value (so far as I can see) for the psychological problems with which I was concerned does not rule out its possible value for a " field-theoretical treatment " of cortical processes[3]. Here my difficulty is that Dr. Wassermann gives no equations of his own to show how his deductions follow from the equations of the quantum physicist; nor does he report any piece of research in which his methods have been actually applied.

[1] W. Ross Ashby, *An Introduction to Cybernetics*, 1956, pp. 97, 158, and 163. Those who dislike the assumptions thus involved may adopt the still more general definitions based on the ' theory of sets '.

[2] C. G. Jung and W. Pauli, *The Interpretation of Nature*, 1955. The principle is not unlike Leibniz's doctrine of pre-established harmony with an empirical rather than a meta-physical basis. I confess, however, that I find Jung's arguments neither clear nor convincing. The re-statement of the principle by Pauli (*op. cit.*, pp. 136 f.)—himself a leading contributor to quantum-theory—seems far more in keeping with the assumptions of modern physics and the requirements of modern psychology.

[3] The gaps at the synaptic junctions are apparently only one or two hundred Ångstrom units across; and therefore, as I have argued elsewhere (e.g. this *Journal*, XI, p. 82), the special conclusions drawn from quantum theory (notably Heisenberg's uncertainty principle) would undoubtedly be relevant to crucial processes taking place in such extremely small regions. I should myself have thought it was far simpler to deal with them in terms of the equations cited in my paper than with those deduced from the quantum theory of fields. Which is the more fruitful methodological device is a question that can only be decided by actual trial.

But for these, no doubt, we must await his promised volume on ' the cybernetics of brain models ' [1].

[1] It is only fair to state that Dr. Wassermann's paper was intended primarily for those interested in paranormal psychology. He claims that clairvoyancy, telepathy, psychokinesis, object-reading, and the like, all " fit well into the framework of theoretical physics ", provided the latter is interpreted so as to include his postulated quantum theory fields; and he believes his theory " avoids the difficulties which C. D. Broad thought would make a purely physical field theory of clairvoyance impossible ". The idea that quantum theory in its more recent forms might have a direct bearing on the problems of paranormal psychology is by no means so original as Dr. Wassermann appears to suppose. Two eminent physicists, Jordan (himself partly responsible for the matrix formulation of quantum theories) and Margenau have themselves put forward such a suggestion : (cf. P. Jordan, ' Atomic Physics and Parapsychology ', *Newsletter of the Parapsychology*, II, iv and refs.; also H. Margenau, ' Physics and Psychic Research ', *ibid.*, III, i and *id.*, ' The nature of Physical Reality, 1950.) More recently Professor C. T. K. Chari has systematically reviewed the whole problem with detailed references (' Quantum Physics and Parapsychology ', *J. Parapsychol.*, XX, pp. 166–183); and has come to the conclusion (but without explicitly mentioning Dr. Wassermann's paper) that " these hypotheses and postulates of modern quantum physics make no *specific* contribution to the problems of parapsychology ". He agrees however that they do have a more general bearing on such problems by " the lesson of restraint offered to those who would reject psi phenomena on physical grounds ". When I myself stated that the majority of psychologists regarded research on telepathy, etc., as unprofitable largely because " there is no known hypothesis which will account for such phenomena and bring them into line with the postulates and principles of modern science ", I too had overlooked Dr. Wassermann's paper; and for that I certainly owe him an apology. Nevertheless, I think I should still have considered his hypothesis as forming no exception if only for the reasons given in my article (footnote 2, p. 73).

BOOK REVIEWS

Theory and Methods of Scaling. By WARREN S. TORGERSON. New York: John Wiley & Sons, Inc. London : Chapman & Hall, Inc., 1958. Pp. xiv+460. 76*s*.

This book was prepared by the author at the request of a committee appointed by the Social Science Research Council in America, and may therefore be considered an authoritative survey of current methods of scaling. After a logical analysis of the different types of measurement, and the role which they play in the development of scientific theory, Dr. Torgerson attempts to introduce some kind of order into the veritable jungle of scaling techniques which has sprung up in recent years. A preliminary discussion of the various criteria which could be adopted for such a purpose leads him to devise a scheme of classification in terms of " the ways the observed variability of responses to stimuli can be allocated." On this basis Torgerson distinguishes three broad approaches.

With the first, the " Subject-Centered Approach ", the systematic variation in the subjects' responses is attributed to individual differences in the subjects themselves. By far the greater part of the work on mental testing is based on measurement of this kind. Most tests of aptitude, achievement, and intelligence are essentially examples of measurement by definition; and in Dr. Torgerson's opinion this approach " has not led, to any great extent, to the development of scaling models ". After this pronouncement it is hardly surprising to find that he excludes this mode of approach from his survey.

The second is the " Stimulus-Centered or Judgement Approach ". Here the systematic variation in the subjects' reactions is attributed to differences in the stimuli with respect to a designated attribute. In successive chapters the author deals in detail with such methods as subjective estimation, fractionation, equisection, differential sensitivity. Three chapters are then devoted to Thurstone's judgement scaling model and the modifications it has undergone, and a further chapter to the technique of multidimensional scaling—a method to which Dr. Torgerson himself has made important contributions.

The third main approach is termed the " Response Method ". Here the variability is ascribed to variations both in the subjects and in the stimuli. This, according to the author's scheme, represents the most general approach possible, and has but recently been developed: a well known instance is Guttman's scalogram technique. Two further chapters deal with various other methods falling under this head, particularly those in which probability forms a fundamental aspect: examples of these techniques are Lazarsfeld's " Latent Structure " models and Coomb's " Unfolding " technique.

All are fully discussed with special reference to the underlying theory, the experimental procedures for obtaining the data, the analytical procedures for estimating the relevant quantities, and the methods by which goodness of fit may be assessed. The exposition is clear and logical; and the clarity is enhanced by the way in which the mathematical derivations of the theoretical properties of each model are kept separate from the mathematics involved in the problems of statistical estimation. It may be safely said that the book will be almost indispensable to any psychological investigator faced with problems of scaling.

A. R. JONCKHEERE.

The Fifth Mental Measurements Yearbook. Edited by Oscar Buros. New Jersey: The Gryphon Press, 1959. Pp. xxviii+1291. $20.

As those who have used the earlier yearbooks will be well aware, the purpose of the series is to aid test users of every kind—teachers, research-workers, educational, industrial, clinical, and statistical psychologists each in his own specific field—to locate and evaluate

mental tests of any given type, and to identify the more important books and articles dealing with each one. It is astonishing that these invaluable volumes are not better known. Almost every day the staff of psychological departments and the editors of psychological journals receive requests from many quarters inquiring about the existence or the merits of tests of various types; and the answer almost invariably is, " You will find all the information you want in the latest *Yearbook* ".

In the present volume, which is larger than any of its predecessors, the section on ' Tests and Reviews ' covers nearly 1,000 tests and 700 test-reviews written specifically for this volume, and numerous excerpts from reviews in various psychological journals. The section on ' Books ' includes nearly 500 volumes with over 500 critical extracts from more than 80 journals. Psychologists on this side of the Atlantic will feel gratified to find that a yearbook coming from a country which has itself contributed so lavishly to test construction is dedicated to a British pioneer—Sir Cyril Burt.

Since a vast amount of the labour, expense, and organization entailed by the selection and arrangement of the material for this volume has been borne personally by Professor and Mrs. Buros, test-users in every country are under an immense debt to them for their courageous and self-sacrificing enterprise. If only in his own interests, every psychologist and educationist should see to it that the series has his fullest support.

CHARLOTTE BANKS.

Planning of Experiments. By D. R. Cox. London: Chapman & Hall. New York : John Wiley & Sons, Inc., 1958. Pp. viii + 308. 60*s*.

Dr. Cox's main purpose is to discuss, as simply as possible, the planning of experiments in those branches of science in which the irrelevant and uncontrolled variations may be of much the same size as those which the investigator desires chiefly to study. Throughout the work he has endeavoured to avoid mathematical technicalities, and to explain the procedures proposed rather on the basis of their intuitive reasonableness than as devices deduced from abstract statistical principles, such as the general theory of analysis of variance. In the earlier part of the book most of the discussion is mainly in terms of concrete examples; and the various illustrative problems have been skilfully selected from a wide range of topics —agriculture, zoology, medicine, psychology, and various branches of industrial and social research. This is just the type of approach suited to the psychological student, who is so often bewildered by the complex algebraic formulae and the mathematical preliminaries that appear in most books on the subject.

The first nine chapters deal with the fundamental concepts and such basic designs as randomized blocks and Latin squares; the last five are concerned with mere advanced techniques such as cross-over designs, fractional replication, and incomplete non-factorial designs. In addition Dr. Cox discusses, in an instructive and comparatively simple way, several topics which are not, as a rule, given adequate prominence in books on experimental design—the reasons for randomization and the practical difficulties involved, the relation of covariance to randomized blocks, the use of contour diagrams for response surfaces, and the various types of factor that may occur in factorial investigations: the attention of the advanced student is even called to the relevance of group theory and Galois's field theory to such problems as those of interaction. Detailed references are appended to each chapter which will enable the research-worker to obtain more detailed information about any specific technique; and a general bibliography on the whole subject is added at the end of the book.

CYRIL BURT.

The Growth of Logical Thinking. By Barbel Inhelder and Jean Piaget. Routledge and Kegan Paul, 1958. Pp. xxvi + 356. 32*s*.

The double title of this work in the original French edition—*De la logique de l'enfant à la logique de l'adolescent: essai sur la construction des structures opératoires formelles*—indicates the double approach adopted by the writers who have collaborated in its production. Professor Piaget tells us how he himself " left his earlier experimental work in order to undertake the study of genetic epistemology and the transformation of logical operations "; at the same time Miss Inhelder embarked on a series of empirical investigations to inquire into " the methods of inductive discovery of physical laws by children and adolescents ". The final outcome has been that both the theorist and the experimenter arrived at very similar and suggestive conclusions by their complementary lines of approach.

The experimental results, we are told, fully corroborate the hypothesis put forward by Piaget in his earlier works. This he describes as " an operational theory of intelligence, aiming at step-by-step explanation of the gradual hierarchical organization of the child's intellectual processes ". There are four main stages, each clearly demarcated from the rest —the ' sensori-motor ', the ' pre-operational ', the stage of ' concrete operations ', and the stage of ' formal operations '. Each successive phase represents the emergence of entirely novel ways of thinking. At the earlier levels " the only logical operations the child can handle are the elementary groupings of classes and relations "; at the highest level, a new kind of reasoning develops—" a propositional logic which involves a combinatorial system of propositional lattices ".

The most interesting sections of the book are the detailed descriptions of the ingenious experiments devised by Miss Inhelder, and the numerous first hand observations she was able to make. For the most part she appears to have limited her studies almost entirely to problems of elementary science—the laws of floating bodies, the angles of incidence and reflection, the oscillation of the pendulum, the conservation of motion, the effects of centrifugal forces, and the like. The method adopted consisted in presenting each problem in concrete form with the aid of simple apparatus to about half a dozen children at various ages, and recording their actual operations and their own explanations of the procedures they invented.

In this general plan, however, there are several obvious shortcomings. To begin with, the rather sweeping conclusions drawn by the writers can hardly be established on the basis of a mere impressionistic analysis of the performances of so small a group: some broader kind of statistical investigation is essential to confirm or check the generalizations so reached. Secondly, experiments involving special apparatus and the discovery of scientific laws must be greatly influenced by the child's previous experience both in and out of school. Many of the explanations offered by the older pupils are couched in technical terms which almost certainly must have been learnt in the classroom. Nor can the ways in which children reason about these rather sophisticated questions be taken as typical of their ordinary modes of reasoning in everyday life.

We are given no data regarding the mental ages of the children. If we may trust the results reported by Burt, Wheeler, and their co-workers in experiments on random samples of the school population with more natural types of problem, it would appear that the child of normal intelligence is fully capable of ' propositional logic ' at a much earlier stage than Professor Piaget and Miss Inhelder have found. As Miss Wheeler points out, the use of this type of reasoning is most likely to occur when the child is arguing with other children, not when he is working by himself or explaining his methods to an older investigator who already knows the answers. Nor has she found any evidence to support the hypothesis of a succession of qualitatively different stages. Here, however, the disagreement is perhaps rather a matter of verbal description and of expository emphasis. Whether or not we accept their final conclusions, the work of Professor Piaget and his colleague is full of instructive points; and here as elsewhere his customary method of reporting verbatim extracts from children's actual conversations is as delightful as it is informative.

R. B. Lewis.

A Psychological Study of Typography. By CYRIL BURT. Cambridge: At the University Press, 1959. Pp. xx+68. 15*s.* net.

The investigations described in this volume originated in a series of researches carried out by Cyril Burt in his capacity of Psychologist to the London County Council and James Kerr, at that time Chief Medical Officer in the Council's Education Department. The immediate object was to determine the style and size of type appropriate to the school books designed for school children of various ages. Later the investigations were extended to cover the printing of books and journals intended for adult readers. With the increasing range of type-faces that have since become available, a repetition of the original inquiries became urgently necessary; and the development of newer experimental and statistical techniques enabled the later investigations to be carried out on a more scientific basis. The main conclusions have already been described in the pages of this *Journal*.

The importance, both of the methods and of the results, seems to have been promptly recognized by many of those professionally engaged in printing and publishing; and the Cambridge University Press accordingly suggested the republication of the work in book form. Dr. Stanley Morison (formerly Reader in Bibliography in the University of Cambridge and himself the designer of one of the most pleasant and legible of contemporary type-faces) has contributed a valuable introduction, reviewing the principal findings from the standpoint of the typographical expert. A long historical appendix describes the early origins and the more conspicuous characteristics of the more important type-faces, and in itself provides a fascinating study. The more technical points are fully illustrated by comparative tables giving specimens of the commonest type-faces. The book will be of special value to teachers, educationists and editors of scientific journals; and will give an added interest to the everyday task of reading books and articles. To psychologists it offers an instructive model for planning future researches on complex practical problems, where the combination of introspective, experimental and statistical devices is essential.

W. L. PARSONS.

The Uses of Argument. By STEPHEN TOULMIN. Cambridge: At the University Press, 1958. Pp. viii+264. 22*s.* 6*d.*

From Aristotle to the present day logic has commonly been regarded as the formal study of rational argument; but the relation of logic as thus interpreted to ordinary everyday reasoning remains obscure. Professor Toulmin starts from an examination of actual procedures in different fields of argument, that is, from the practice of logic rather than its theory, and thus finds a far richer variety of processes than is envisaged in any purely formal system. Other logicians treat logic as the *scientia scientiarum*—the science of scientific reasoning, with mathematics as the ideal; Professor Toulmin would prefer to take the legal arguments of the jurist as the typical model.

The psychologist may be tempted to maintain that the study of actual precedures and of everyday methods of disputation belongs not so much to logic as to the psychology of thought-processes, or, as it is commonly called, ' problem-solving '. Be that as it may, those who undertake to investigate the development of reasoning in the child or the race, or to analyse its manifestations in the average educated or uneducated adult, will find Professor Toulmin's chapters full of stimulating questions and suggestive hypotheses, which well deserve systematic study from the empirical standpoint of the psychologist.

W. L. GREGORY.

Human Potentialities. By GARDNER MURPHY. New York: Basic Books Inc., 1958. Pp. x+340. $6.00.

Professor Murphy's new book forms (as he himself puts it) a bold attempt at " long-range weather-forecasting for humanity ". We are living through a stormy period: and meteorologists have learnt much by studying storms, depressions, cyclonic disturbances, and the menaces of cold, hot, and occluded fronts. His aim is to discuss how the facts, methods, and limitations characteristic of modern science point the way to the future direction of human development. Viewing man as a still-evolving organism, he seeks to show how man himself, by exercising his own intelligence and by an effort of conscious choice, can transcend both his biological and his cultural inheritance, and so " open up a way for the free expression of his virtually unlimited potentialities ".

The book is divided into five main parts. The first three are concerned with what Dr. Murphy calls the ' three human natures '. The ' first human nature '—the raw basis —is determined partly by the general biochemical and neurological organization common to all human beings, and partly by the individual diversity superimposed on this by each man's unique genetic constitution. The ' second human nature ' he terms " the mould of culture." In spite of the amazing advances resulting from individual development in a creative social environment, there has been, he believes, a tendency for the ' second human nature ' to harden and gradually become fixed in a rigid system which is ill adapted to the needs of a swiftly changing world. The ' third human nature ' is or should be one which " frees intelligence and breaks through the mould ". The brilliant achievements of quantitative techniques have led to grave neglect of the qualitative aspects. Such novel discoveries as those of quantum theory and parapsychology present a challenge to the human mind which neither the ordinary scientist nor the ordinary philosopher has so far been able to meet. And accordingly the concluding chapters of the book attempt to indicate " how the three human natures can be expected to coalesce and integrate, and how the various attributes of man-to-man and man-to-cosmos patterns can be expected to emerge ".

T. R. WATSON.

The Evolution of Genetic Systems. By C. D. DARLINGTON. Edinburgh: Oliver & Boyd, 1958. Pp. xii+265. 21*s*.

This is a new edition, brought up to date and much enlarged, of a book first published twenty years ago. Its purpose is " to show how helpful genetics can be to other sciences, and how helpful other sciences can be to genetics ". Even today, as the author rightly reminds us, the full implications of Mendel's work are still largely overlooked and widely misunderstood. The popular formulation of it in terms of would-be scientific laws—a ' law of segregation ' and a ' law of recombination '—is grossly inaccurate. Rightly interpreted, " it makes the character of the individual, not predictable, but inherently unpredictable ". Hence the need for statistical treatment (as Mendel was the first to realize); and hence too the obvious limitations of any such treatment.

In Dr. Darlington's opinion it is " the union of a material basis for heredity with a general evolutionary framework " which provides the true method, and indeed the only method, of making sense of biology as a whole. The entire complicated process, so far as it is now understood, is set forth in these pages in a lucid, detailed, and systematic fashion; and, since psychology is itself a biological science, this authoritative survey is one with which every psychologist should make himself familiar.

J. CONWAY.

The Sleep Walkers. By ARTHUR KOESTLER. London : Hutchinson, 1959. Pp. 624, 25*s*.

This is one of the most instructive of the many volumes that have come from the versatile pen of Mr. Koestler. The book itself is essentially a psychological study of the background, the personalities, the working methods, and the basic modes of thinking exhibited by the four great pioneers who paved the way for modern science—Copernicus, Kepler, Brahe, and Galileo. For this purpose he has relied largely on the correspondence of each of his subjects; and not the least fascinating parts of the book are the numerous letters and other documents, quoted verbatim and often in full. Thus the whole work is, like Mr. Koestler's previous volume on *Insight and Outlook*, a study of what he calls " the psychological process of discovery ".

As Professor Butterfield points out in his introduction, Koestler corrects many traditional misinterpretations and " gives the whole subject a number of unexpected ramifications; . . . we have read too much modernity into Copernicus, selected things with a modern ring from Kepler, and been wholly anachronistic in our treatment of Galileo ". To a large extent, therefore, he breaks down the popular myth of the rational, logical, methodical advance of science. Scientists themselves, with their restricted outlook and their automatic manipulations of simple postulates and concepts, appear in his eyes rather as " sleepwalkers," groping painfully through a fog of muddled ideas and residual beliefs into a twilight that may, if they awake, herald a new dawn. Having studied science at Vienna, when Vienna itself was helping to initiate a new epoch in science and logic, Koestler is admirably equipped for the purpose. As a result the importance of Copernicus and Galileo is noticeably diminished, that of Brahe and Kepler—the first investigators to appreciate the need for quantitative measurement and mathematical techniques—is greatly enhanced. Mediaeval religion, with its firm belief in the intelligible rationality and lawfulness of the universe, appears (as Whitehead had already maintained) more as an active supporter than as an obstructive opponent of the scientific renaissance. Physicists and philosophers may both be tempted to question some of Mr. Koestler's own conclusions; but the psychologist at any rate can learn much from his brilliant studies of the psychology of genius. W. L. GREGORY.

The Sixth Sense: An Inquiry into Extra-Sensory Perception. By ROSALIND HEYWOOD. London : Chatto & Windus, 1959. Pp. 96. 21*s*.

Mrs. Heywood's book provides a very readable survey, at once sympathetic and critical, of " the whole family of apparent phenomena and experiences, real or alleged, for which no physical cause has as yet been discovered ", grouped under the heading of 'psychical research'. For this she is unusually well equipped. She herself has had experiences that might be classed as ' paranormal ', though she deprecates the term. She has published investigations of her own; and has enjoyed a long and intimate acquaintance with the work of the Society for Psychical Research.

The problems involved have been attacked along two complementary lines: the earlier investigators confined themselves chiefly to examining alleged phenomena which appeared sporadically and spontaneously; later investigators have relied rather on experimental data, obtained under carefully controlled conditions such as others could repeat, and subjected to a rigorous statistical analysis. And, in discussing the relations between the two methods, Mrs. Heywood brings out—what the statistical psychologist is rather apt to forget in this as in other fields—namely, that the artificial problems which he investigates are, as a rule, so far removed from everyday life that they only gain meaning when his clear-cut quantitative conclusions are re-interpreted in the light of the vaguer qualitative observations which have been made incidentally and under more natural conditions.

After a brief historical review, Mrs. Heywood begins with the question which mainly attracted the original pioneers—the survival of a psychic entity or factor after physical death. Apart from the rather doubtful evidence provided by ' cross-correspondences ', she concludes that clairvoyance, telepathic interaction, or even sheer coincidence could be made to account for nearly all the phenomena reported. She then turns to experimental studies. She describes first the more striking results provided by experiments that do not lend themselves to statistical analysis. Of these by far the most convincing are the results reported by Gilbert Murray, who himself displayed what were apparently unusual telepathic powers. Murray is the last person who could be suspected of exhibitionism, wishful thinking, or conscious or unconscious fraud. He originally supposed that he must be unintentionally exploiting an exceptional gift of auditory hyperaesthesia. But later experiments completely ruled out any explanation of this type. His introspective reports are particularly suggestive. The experience, he says " does not feel quite like cognition or detection, but more like the original sense of the word ' sympathy ': I seem to be passive, and to feel in a shadowy way the state of the mind of someone else ". And he adds: " I never had any success in guessing cards or numbers, or any subject that was not in some way interesting or amusing ".

Mrs. Heywood then devotes two chapters to a summary of the results of the more important quantitative experiments. As long ago as 1885 Sir Oliver Lodge suggested that the most cogent way to demonstrate telepathy or clairvoyance would be to use card-guessing, and proposed a formula for testing the statistical significance of the results. The attempts of Coover, of Miss Jephson (with statistical assistance from Sir Ronald Fisher), of Soal, and of Rhine are described in some detail. In spite of the objections raised by Mr. Spencer Brown, Mrs. Heywood herself concludes that " the best available method of confirming ESP lies in quantitative experiments of these various types ".

Nevertheless, the mere fact that a student can on an average guess 8 cards out of 25 (instead of the 5 one would expect), and that the difference is statistically significant, is of little or no interest in itself. If, however, on the basis of card-guessing we can produce strong prima facie evidence in favour of telepathy or clairvoyance in trivial matters, then we are on firmer ground in accepting the less conclusive evidence provided by spontaneous manifestations. The concentration of recent research workers almost exclusively on statistical enquiries has, she believes, led most psychologists to overlook the striking observations reported by earlier pioneers and buried in the archives of the S.P.R. In an instructive appendix she summarizes the views of half a dozen eminent authorities who have accepted at least some of the evidence and tried to suggest explanatory hypotheses—physicists like G. N. M. Tyrrell, biologists like Sir Alister Hardy, psychiatrists like Jung and Freud, psychologists like Thouless and Gardner Murphy, and (perhaps most instructive of all) philosophers like H. H. Price and C. D. Broad.

T. A. Harris.

INDEX TO VOLUME XII

ARTICLES

Author	*Title*	*Page*
Brown, J., and Cane, V. R.	An Analysis of the Limiting Method	119
Burt, C.	Class Differences in Intelligence, III	15
Burt, C.	Experiments on Telepathy in Children	55
Burt, C.	Field Theories and Statistical Psychology	153
Cane, V. R., and Brown, J.	An Analysis of the Limiting Method	119
Conway, J.	Class Differences in Intelligence, II	5
Gregory, C. C. L.	Statistical Methods in Psychical Research, I	139
Halsey, A. H.	Class Differences in Intelligence, I	1
Highmore, G., and Jones, B. W.	The Athletic Ability of Boys in a Secondary Modern School	127
Jones, B. W., and Highmore, G.	The Athletic Ability of Boys in a Secondary Modern School	127
Peel, E. A.	The Measurement of Interests by Verbal Methods	105
Sandon, F.	Twins in the School Population	133
Soal, S. G.	Statistical Methods in Psychical Research, II	149
Wrigley, C.	The Effect upon the Communalities of Changing the Estimate of the Number of Factors	35

BOOK REVIEWS

Author	*Title*	*Page*
Bartlett, F. C.	Thinking: An Experimental and Social Study	102
Burnett, W.	This is My Philosophy	103
Buros, O.	The Fifth Mental Measurement Yearbook	165
Burt, C.	A Psychological Study of Typography	168
Cox, D. R.	The Planning of Experiments	166
Darlington, C. D.	The Evolution of Genetic Systems	169
Garrett, H. E.	Statistics in Psychology and Education	101
Heywood, R.	The Sixth Sense: An Inquiry into Extra-Sensory Perception	170
Himmelweit, H. *et al.*	Television and the Child	104
Inhelder, B., and Piaget, J.	The Growth of Logical Thinking	167
Koestler, A.	The Sleepwalkers	170
Macpherson, J. S.	Eleven-Year-Olds Grow Up	101
Murphy, G.	Human Potentialities	169
Schroedinger, E.	Science, Theory, and Man	103
Torgerson, W. S.	Theory and Methods of Scaling	165
Toulmin, S.	The Uses of Argument	168

Vol. XII Part I May 1959

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
CYRIL BURT AND JOHN WHITFIELD

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ALAN STUART
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
P. E. VERNON

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s. per part* (*U.S.A.* $3.50) *Subscription* 30*s.* 6*d. per volume* (*U.S.A* $4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s.* net for Members of the Society and 30*s.* 6*d.* (post free) for non-members. The subscription price in the U.S.A. is $4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1 : non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, in May and November.

Papers for publication should be sent to JOHN WHITFIELD, University College, Gower Street, London, W.C.1. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 25 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

Books for review should be sent to the Editor, advertisements to the Secretary, British Psychological Society.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology* : subscriptions, 60*s.* per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* : for this the subscription, £1 per annum, should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society.

Papers for publication in *The British Journal of Psychology* should be sent to Professor James Drever, Department of Psychology, the University, Edinburgh; those for publication in *The British Journal of Medical Psychology* to Dr. Joseph Sandler, 96 Portland Place, London, W.1; those for publication in *The British Journal of Educational Psychology* to Professor P. E. Vernon, London University Institute of Education, Malet Street, London, W.C.1.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract : (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the ' Recommendations to Authors ' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on ' Setting Mathematics ' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost.

In future authors will receive two complete slip-proofs of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

OCCUPATIONAL PSYCHOLOGY

Editor : ALEC RODGER

APRIL, 1959 VOLUME 33, No 2.

An Autobiographical Sketch — MARY BOOLE STOTT

Time Relationships in Pilot Performance — K. F. JACKSON

Fundamentalism in Industrial Psychology — MORRIS S. VITELES

The Utility of Consumer Motivation Research — R. A. M. GREGSON

Book Reviews

Other Books Received

Annual Subscription **40** *shillings*

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

"GRIFFIN'S STATISTICAL MONOGRAPHS & COURSES"

Edited by M. G. KENDALL, Sc.D.

A new series of publications planned to meet the need for a form of book, at moderate cost, which will make accessible, to a group of readers, specialized studies in statistics or special courses on particular statistical topics.

No. 1. **ANALYSIS OF MULTIPLE TIME-SERIES**
M. H. QUENOUILLE — 24s. net

No. 2. **A COURSE IN MULTIVARIATE ANALYSIS**
M. G. KENDALL — 22s. net

No. 3. **THE FUNDAMENTALS OF STATISTICAL REASONING**
M. H. QUENOUILLE — *In the press*

THE MATHEMATICAL THEORY OF EPIDEMICS

By NORMAN T. J. BAILEY, M.A.
Reader in Biometry, University of Oxford.

A valuable survey, with the main emphasis on the biometrical and epidemiological rather than the mathematical aspects.

General introduction—Historical sketch of mathematical epidemiology—Epidemiological principles—Deterministic theory—Stochastic theory : continuous-infection models—Stochastic theory : chain-binomial models—The measurement of latent and infectious periods—Recurrent epidemics and endemicity—The detection of infectiousness—Summary and conclusions—Index. **36s. net**

CHARLES GRIFFIN & COMPANY LIMITED
42 DRURY LANE LONDON WC2

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XII Part I May 1959

CONTENTS

		Page
HALSEY, A. H.	Class Differences in General Intelligence, I . .	1
CONWAY, J.	Class Differences in General Intelligence, II . .	5
BURT, C.	Class Differences in General Intelligence, III . .	15
WRIGLEY, C.	The Effect upon the Communalities of Changing the Estimate of the Number of Factors	35
BURT, C.	Experiments on Telepathy in Children (Critical Notice)	55
BOOK REVIEWS	. . . . .	101

LIBRARY
16 OCT 1959
UNIVERSITY OF LONDON

JOURNAL COMMITTEE

CYRIL BURT *and* JOHN WHITFIELD *Editors*

CHARLOTTE BANKS *and* ALAN STUART *Assistant Editors*

and

D. E. BROADBENT
M. HAMILTON
A. N. OPPENHEIM
J. S. SMALL
J. SUTHERLAND
F. W. WARBURTON